US006889216B2

(12) **United States Patent**
Nugent

(10) **Patent No.: US 6,889,216 B2**
(45) **Date of Patent: May 3, 2005**

(54) **PHYSICAL NEURAL NETWORK DESIGN INCORPORATING NANOTECHNOLOGY**

(75) Inventor: **Alex Nugent**, Santa Fe, NM (US)

(73) Assignee: **Knowm Tech, LLC**, Albuquerque, NM (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 585 days.

(21) Appl. No.: **10/095,273**

(22) Filed: **Mar. 12, 2002**

(65) **Prior Publication Data**

US 2003/0177450 A1 Sep. 18, 2003

(51) **Int. Cl.**[7] **G06F 15/18**
(52) **U.S. Cl.** **706/15**
(58) **Field of Search** 706/15

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,802,951	A	2/1989	Clark et al.	156/630
4,974,146	A	11/1990	Works et al.	364/200
4,988,891	A	1/1991	Mashiko	307/201
5,315,162	A	5/1994	McHardy et al.	307/201
5,422,983	A	6/1995	Castelaz et al.	395/24
5,475,794	A	12/1995	Mashiko	395/24
5,589,692	A	12/1996	Reed	257/23
5,649,063	A	7/1997	Bose	395/22
5,706,404	A	1/1998	Colak	395/24
5,717,832	A	2/1998	Steimle et al.	395/24
5,783,840	A	7/1998	Randall et al.	257/24
5,812,993	A	9/1998	Ginosar et al.	706/26
5,904,545	A	5/1999	Smith et al.	438/455
5,951,881	A	9/1999	Rogers et al.	216/41
5,978,782	A	11/1999	Neely	706/16
6,026,358	A	2/2000	Tomabechi	704/232
6,128,214	A	10/2000	Kuekes et al.	365/151
6,248,529	B1	6/2001	Connolly	435/6
6,256,767	B1	7/2001	Kuekes et al.	716/9
6,282,530	B1	8/2001	Huang	706/41
6,294,450	B1	9/2001	Chen et al.	438/597
6,314,019	B1	11/2001	Kuekes et al.	365/151
6,330,553	B1	12/2001	Uchikawa et al.	706/2
6,339,227	B1	1/2002	Ellenbogen	257/40
2001/0004471	A1	6/2001	Zhang	427/372.2
2001/0023986	A1	9/2001	Mancevski	257/741
2001/0024633	A1	9/2001	Lee et al.	423/447.3
2001/0044114	A1	11/2001	Connolly	435/6
2002/0001905	A1	1/2002	Choi et al.	438/268

FOREIGN PATENT DOCUMENTS

EP	1 022 764 A1	1/2000
EP	1 046 613 A2	4/2000
EP	1 100 106 A2	5/2001
EP	1 069 206 A3	7/2001
EP	1 115 135 A1	7/2001
EP	1 134 304 A2	9/2001
WO	WO 00/44094	7/2000

OTHER PUBLICATIONS

Schoenbach et al, "Bioelectrics–New Applications for Pulsed Power Technology", IEEE Digest of Technical Papers on Pulsed Power Plasma Science, Jun. 2001.*

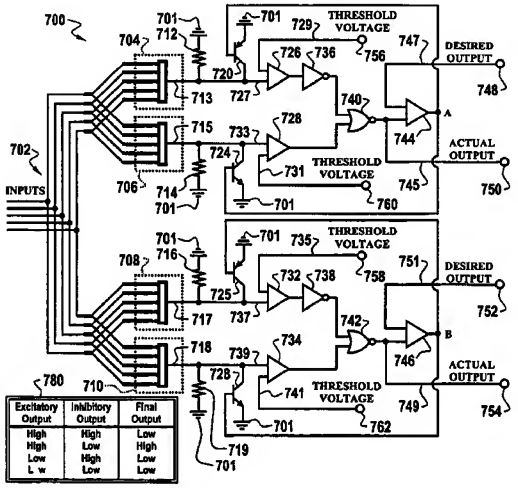
(Continued)

Primary Examiner—George Davis
(74) *Attorney, Agent, or Firm*—Kermit D. Lopez; Luis M. Ortiz; Ortiz & Lopez, PLLC

(57) **ABSTRACT**

A physical neural network based on nanotechnology, including methods thereof. Such a physical neural network generally includes one or more neuron-like nodes, which are formed from a plurality of interconnected nanoconnections formed from nanoconductors. Each neuron-like node sums one or more input signals and generates one or more output signals based on a threshold associated with the input signal. The physical neural network also includes a connection network formed from the interconnected nanoconnections, such that the interconnected nanoconnections used thereof by one or more of the neuron-like nodes are strengthened or weakened according to an application of an electric field.

52 Claims, 6 Drawing Sheets



OTHER PUBLICATIONS

PCT Notification of Transmittal of the International Search Report or the Declaration, Date of Mailing Jun. 4, 2004.
Peter Weiss, "Circuitry in a Nanowire: Novel Growth Method May Transform Chips," Science News Online, vol. 161, No. 6; Feb. 9, 2002.
Press Release, "Nanowire–based electronics and optics comes one step closer," Eureka Alert, American Chemical Society; Feb. 1, 2002.
Weeks et al., "High–pressure nanolithography using low–energy electrons from a scanning tunneling microscope," Institute of Physics Publishing, Nanotechnology 13 (2002), pp. 38–42; Dec. 12, 2001.
CMP Cientifica, "Nanotech: the tiny revolution"; CMP Cientifica, Nov. 2001.
Diehl, et al., "Self–Assembled, Deterministic Carbon Nanotube Wiring Networks," Angew. Chem. Int. Ed. 2002, 41, No. 2; Received Oct. 22, 2001.
G. Pirio, et al., "Fabrication and electrical characteristics of carbon nanotube field emission microcathodes with an integrated gate electrode," Institute of Physics Publishing, Nanotechnology 13 (2002), pp. 1–4, Oct. 2, 2001.
Leslie Smith, "An Introduction to Neural Networks," Center for Cognitive and Computational Neuroscience, Dept. of Computing & Mathematics, Unversity of Stirling, Oct. 25, 1996; http//www.cs.stir.ac.uk/~lss/NNIntro/InvSlides.html.
V. Derycke et al., "Carbon Nanotube Inter– and Intramolecular Logic Gates," American Chemical Society, Nano Letters, XXXX, vol. 0, No. 0, A–D.
Mark K. Anderson, "Mega Steps Toward the Nanochip," Wired News, Apr. 27, 2001.
Collins et al., "Engineering Carbon Nanotubes and Nanotube Circuits Using Electrical Breakdown," Science, vol. 292, pp. 706–709, Apr. 27, 2001.
Landman et al., "Metal–Semiconductor Nanocontacts: Silicon Nanowires," Physcial Review Letters, vol. 85, No. 9, Aug. 28, 2000.
John G. Spooner, "Tiny tubes mean big chip advances," Cnet News.com, Tech News First, Apr. 26, 2001.
Jeong–Mi Moon et al., "High–Yield Purification Process of Singewalled Carbon Nanotubes," J. Phys. Chem. B 2001, 105, pp. 5677–5681.
"A New Class of Nanostructure: Semiconducting Nanobelts Offer Potential for Nanosensors and Nanoelectronics," Mar. 12, 2001, http://www.sciencedaily.com/releases/2001/03/010309080953.htm.
Hermanson et al., "Dielectrophoretic Assembly of Electrically Functional Microwires from Nanoparticles Suspensions," Materials Science, vol. 294, No. 5544, Issue of Nov 2, 2001, pp. 1082–1086.
Press Release, "Toshiba Demonstrates Operation of Single–Electron Transistor Circuit at Room Temperature," Toshiba, Jan. 10, 2001.
J. Appenzeller et al., "Optimized contact configuration for the study of transport phenomena in ropes of single–wall carbon nanotubes," Applied Physics Letters, vol. 78, No. 21, pp. 3313–3315, May 21, 2001.
David Rotman, "Molecular Memory, Replacing silicon with organic molecules could mean tiny supercomputers," Technology Review, May 2001, p. 46.
Westervelt et al., "Molecular Electronics,"NSF Functional Nanostructures Grant 9871810, NSF Partnership in Nanotechnology Conference, Jan. 29–30, 2001; http://www.unix.oit.umass.edu/~nano/NewFiles/FN19_Harvard.pdf.
Niyogi et al., "Chromatographic Purification of Soluble Single–Walled Carbon Nanotubes (s–SWNTs)," J. Am. Chem. Soc 2001, 123, pp. 733–734, Received Jul. 10, 2000.
Duan et al., "Indium phosphide nanowires as building blocks for nanoscale electronic and optoelectronic devices," Nature, vol. 409, Jan. 4, 2001, pp. 67–69.
Paulson, et al., "Tunable Resistance of a Carbon Nanotube–Graphite Interface," Science, vol. 290, Dec. 1, 2000, pp. 1742–1744.
Wei et al., "Reliability and current carrying capacity of carbon nanotubes," Applied Physics Letters, vol. 79, No. 8, Aug. 20, 2001, pp. 1172–1174.
Collins et al., "Nanotubes for Electronics," Scientific American, Dec. 2000, pp. 62–69.
Avouris et al., "Carbon nanotubes: nanomechanics, manipulation, and electronic devices," Applied Surface Science 141 (1999), pp.201–209.
Smith et al., "Electric–field assisted assembly and aligment of metallic nanowires," Applied Physics Letters, vol. 77, No. 9, Aug. 28, 2000, pp. 1399–1401.
Hone et al., "Electrical and thermal transport properties of magnetically aligned single wall carbon nanotube films," Applied Physics Letters, vol. 77, No. 5, Jul. 31, 2000, pp. 666–668.
Smith et al., "Structural anisotropy of magnetically aligned single wall carbon nanotube films," Applied Physics Letters, vol. 77, No. 5, Jul. 31, 2000, pp. 663–665.
Andriotis et al., "Various bonding configurations of transition–metal atoms on carbon nanotubes: Their effect on contact resistance," Applied Physics Letters, vol. 76, No. 26, Jun. 26, 2000, pp. 3890–3892.
Chen et al., "Aligning single–wall carbon nanotubes with an alternating–current electric field," Applied Physics Letters, vol. 78, No. 23, Jun. 4, 2001, pp. 3714–3716.
Bezryadin et al., "Self–assembled chains of graphitized carbon nanoparticles," Applied Physics Letters, vol. 74, No. 18, May 3, 1999, pp. 2699–2701.
Bezryadin et al., "Evolution of avalanche conducting states in electrorheological liquids," Physical Review E, vol. 59, No. 6, Jun. 1999, pp. 6896–6901.
Liu et al., "Fullerene Pipes," Science, vol. 280, May 22, 1998, pp. 1253–1255.
Yamamoto et al., "Orientation and purification of carbon nanotubes using ac electrophoresis," J. Phys. D: Appl. Phys 31 (1998) L34–L36.
Bandow et al., "Purification of Single–Wall Carbon Nanotubes by Microfiltration," J. Phys. Chem. B 1997, 101, pp. 8839–8842.
Tohji et al., "Purifying single walled nanotubes," Nature, vol. 383, Oct. 24, 1996, p. 679.
Dejan Rakovic, "Hierarchical Neural Networks and Brainwaves: Towards a Theory of Consciousness," Brain & Consciousness: Proc. ECPD Workshop (ECPD, Belgrade, 1997), pp. 189–204.
Dave Anderson & George McNeill, "Artifical Neural Networks Technology," A DACS (Data & Analysis Center for Software) State–of–the Art Report, Contract No. F30602–89–C–0082, ELIN: A011, Rome Laboratory RL/C3C, Griffiss Air Force Base, New York, Aug. 20, 1992.
Greg Mitchell, "Sub–50 nm Device Fabrication Strategies," Project No. 890–00, Cornell Nanofabrication Facility, Electronics—p. 90–91, National Nanofabrication Users Network.

John–William DeClaris, "An Introduction to Neural Networks," http://www.ee.umd.edu/medlab/neural/nnl.html.
"Neural Networks," StatSoft, Inc., http://www.statsoftinc.com/textbook/stevnet.html.
Stephen Jones, "Neural Networks and the Computation Brain or Maters relating to Artificial Intelligence," The Brain Project, http://www.culture.com.au/brain_proj/neur_net.htm.
David W. Clark, "An Introduction to Neural Networks"; http://members.home.net/neuralnet/introtonn/index.htm.
"A Basic Introduction to Neural Networks"; http://blizzard.gis.uiuc.edu/htmldocs/Neural/neural.html.
Meyer et al., "Computational neural networks: a general purpose tool for nanotechnology," http://www.foresight.org/Conferences/MNT05/Abstracts/Meveabst.html.
Saito et al., "A 1M Synapse Self–Learning Digital Neural Network Chip," ISSCC, pp. 6.5–1 to 6.5–10, IEEE 1998.
Espejo, et al., "A 16×16 Cellular Neural Network Chip for Connected Component Detection," Jun. 30 1999; http://www.imse.cnm.csic.es/Chipcat/espeio/chip–2.pdf.
Pati et al., "Neural Networks for Tactile Perception," Systems Research Center and Dept. of Electrical Engineering, University of Maryland and U.S. Naval Research Laboratory. 1987; http://www.isr.umd.edu/TechReports/ISR/1987/TR_87–123/TR_87–123.phtml.
Osamu Fujita, "Statistical estimation of the number of hidden units for feedforward neural networks," Neural Networks 11 (1998), pp. 851–859.

* cited by examiner

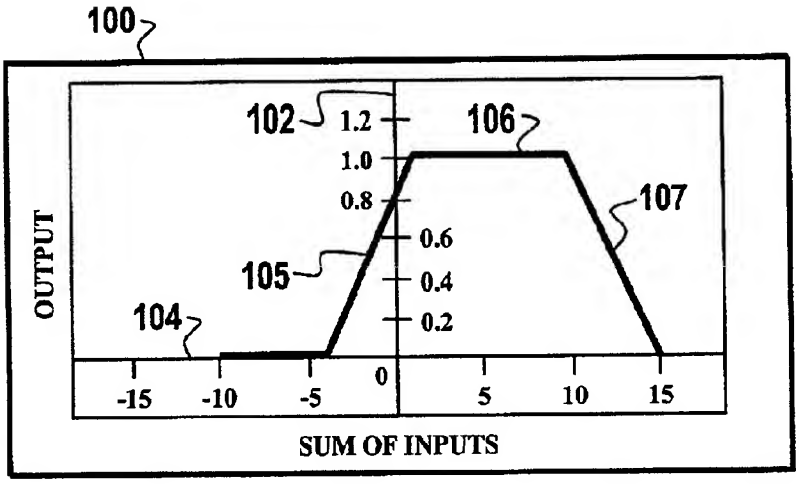


Figure 1

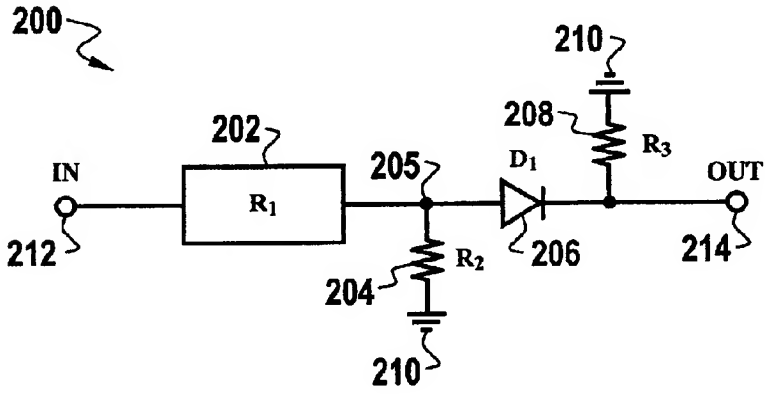


Figure 2

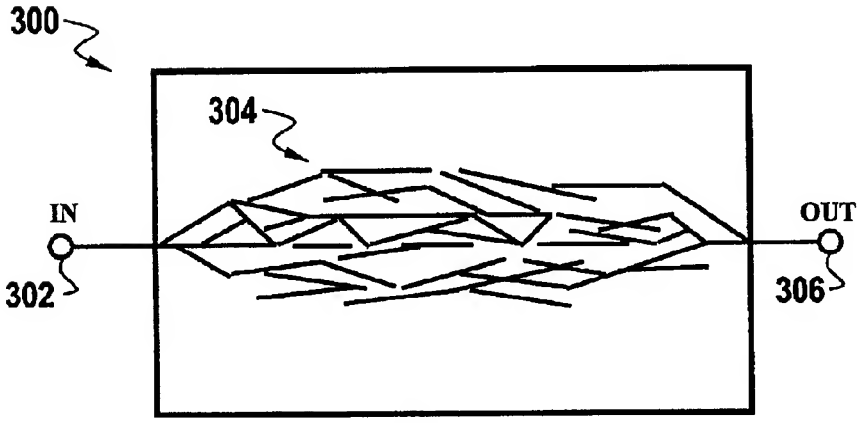


Figure 3

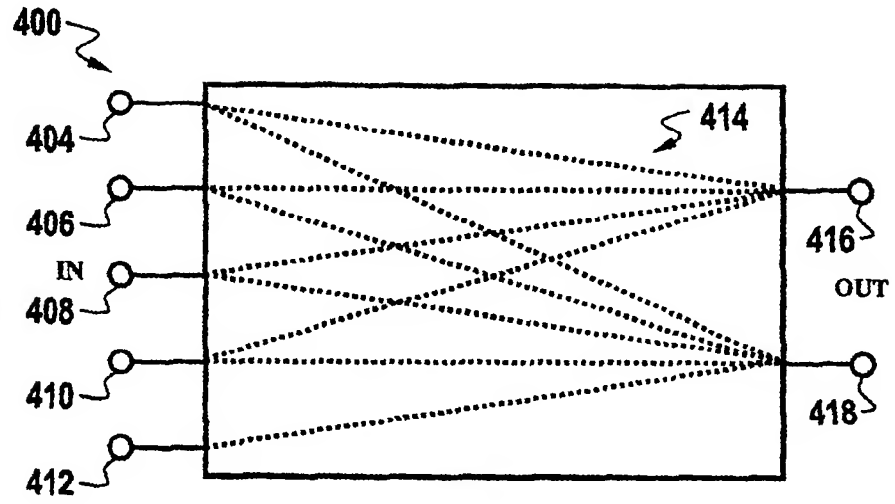


Figure 4

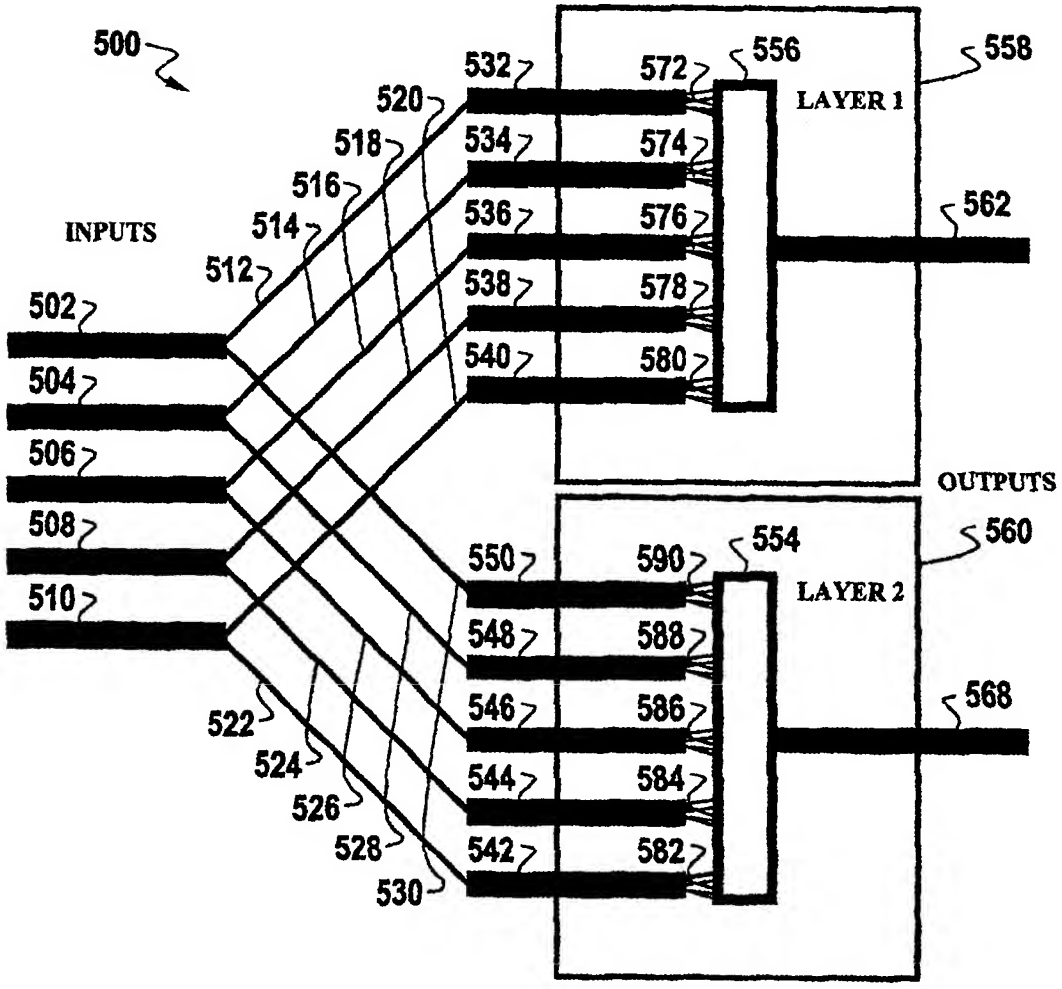


Figure 5

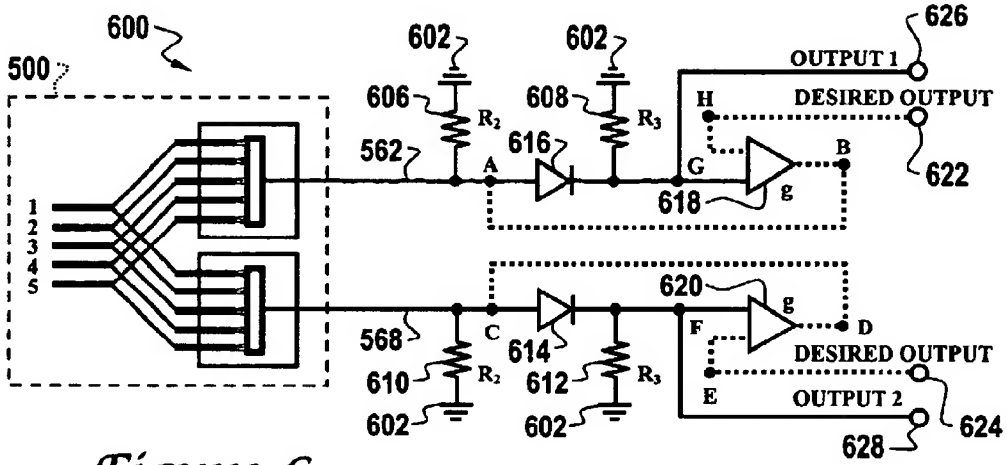


Figure 6

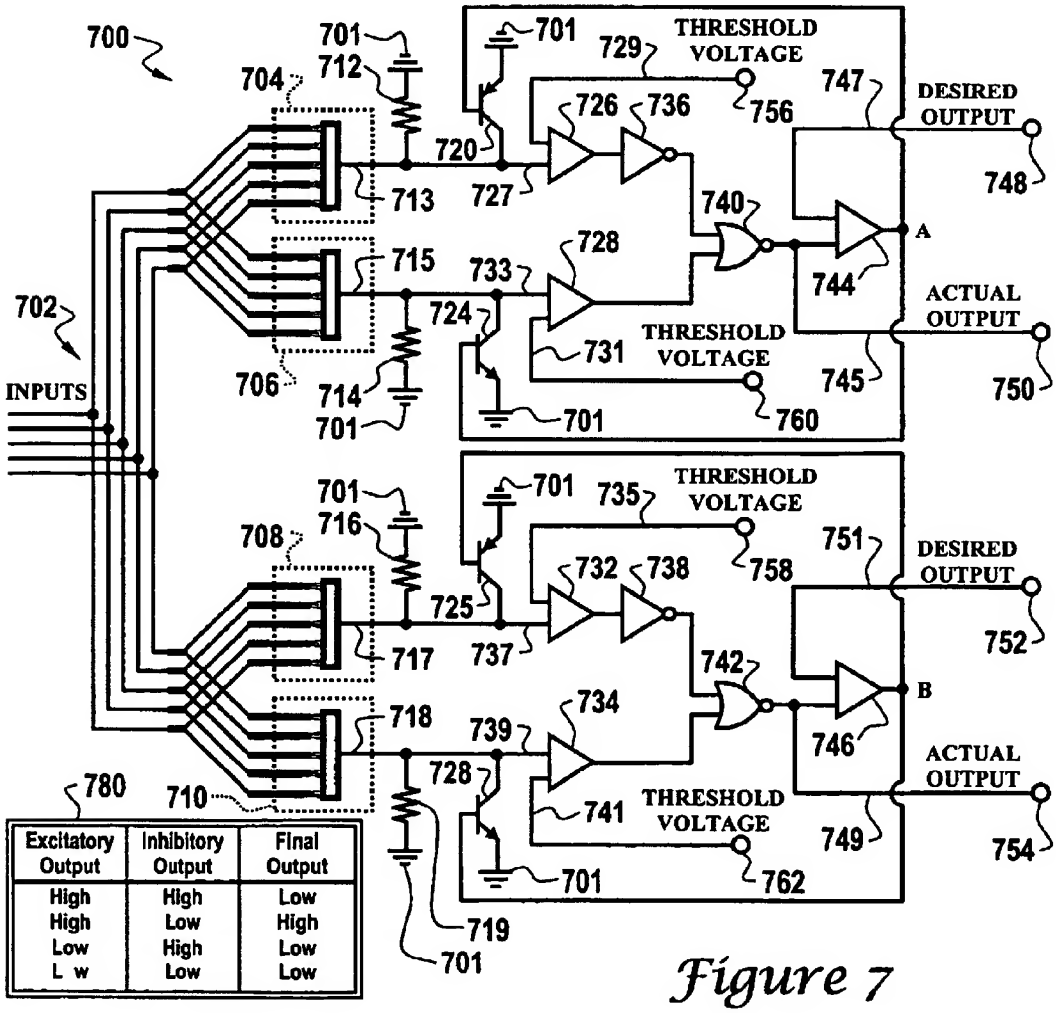


Figure 7

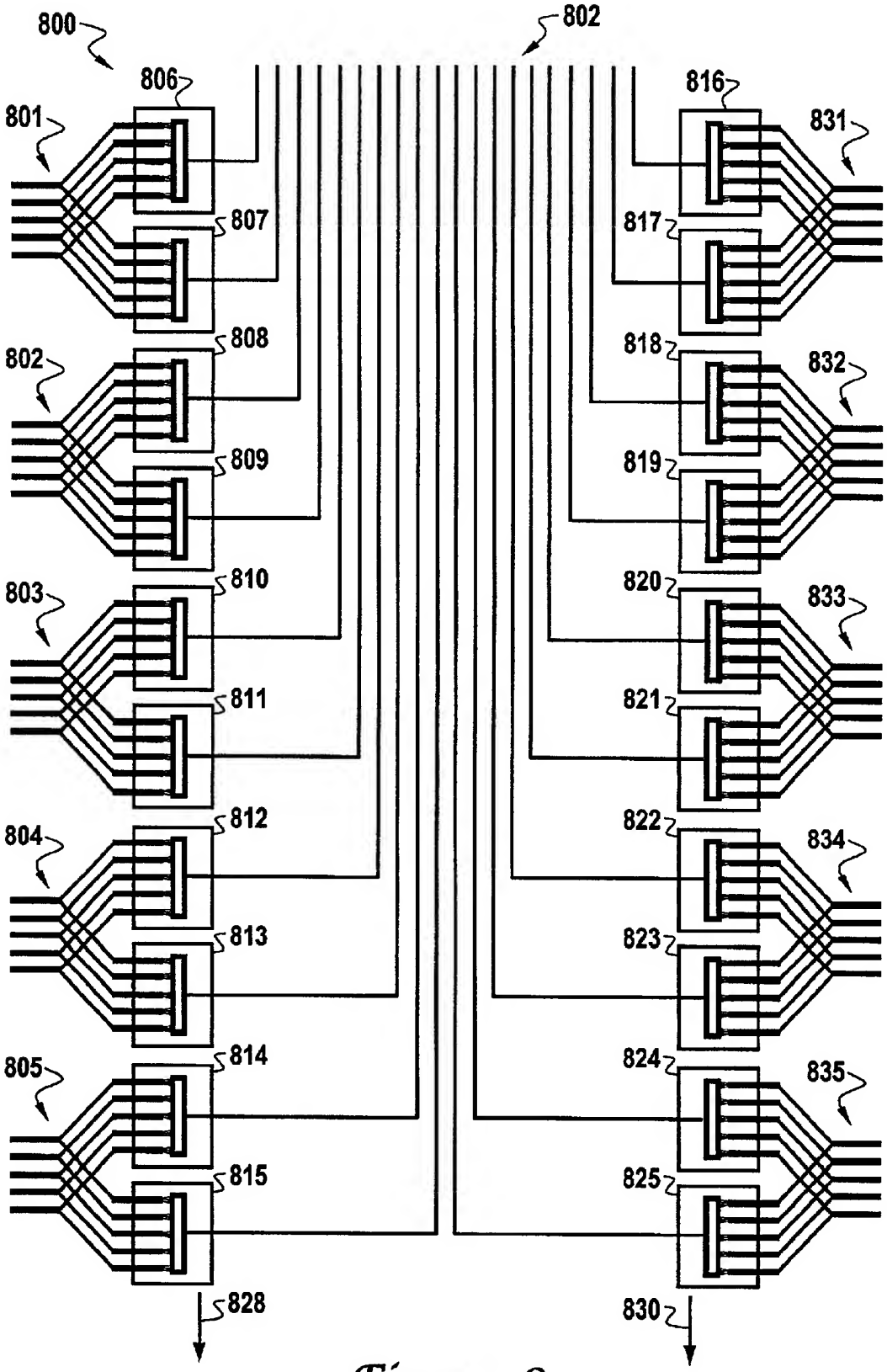

800
802
806
816
801
831
807
817
808
818
802
832
809
819
810
820
803
833
811
821
812
822
804
834
813
823
814
824
805
835
815
825
828
830

Figure 8

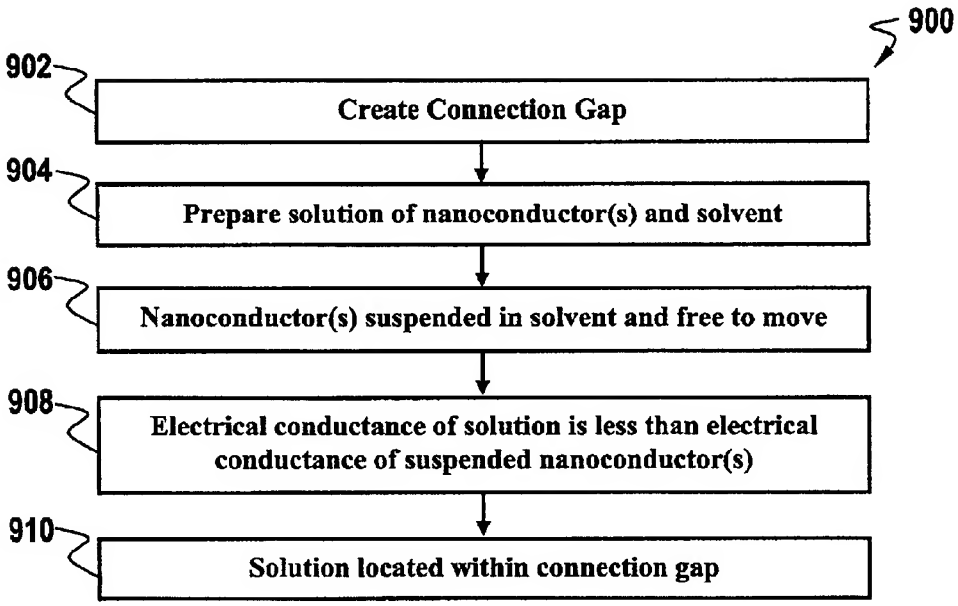


Figure 9

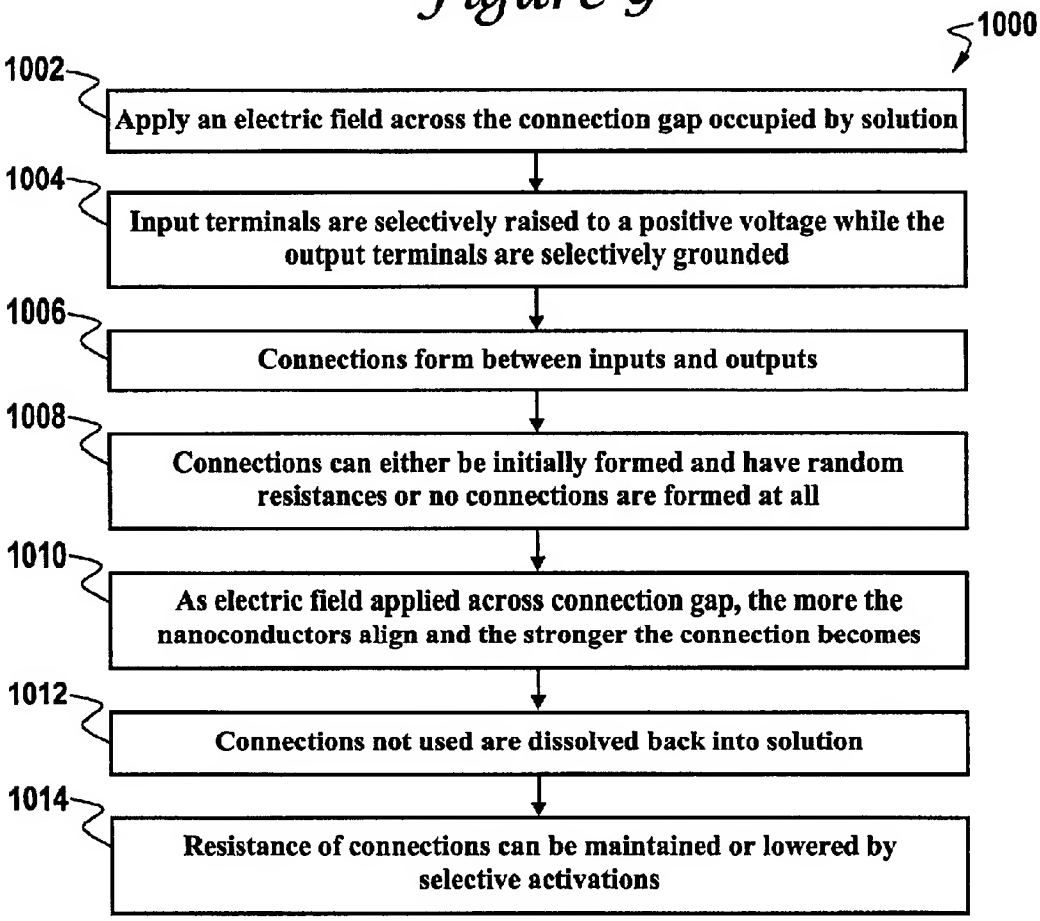


Figure 10

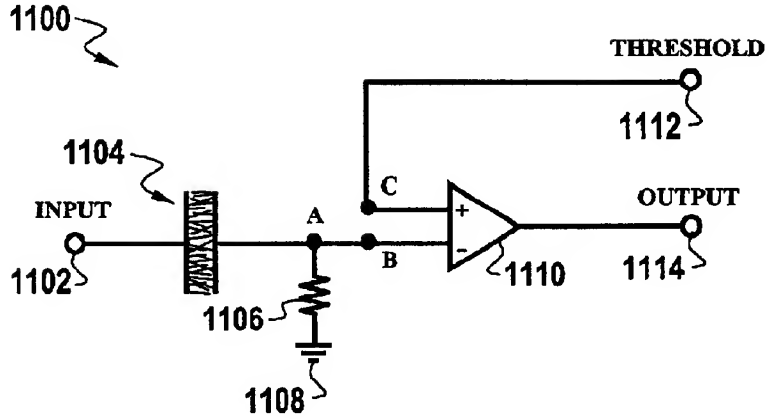


Figure 11

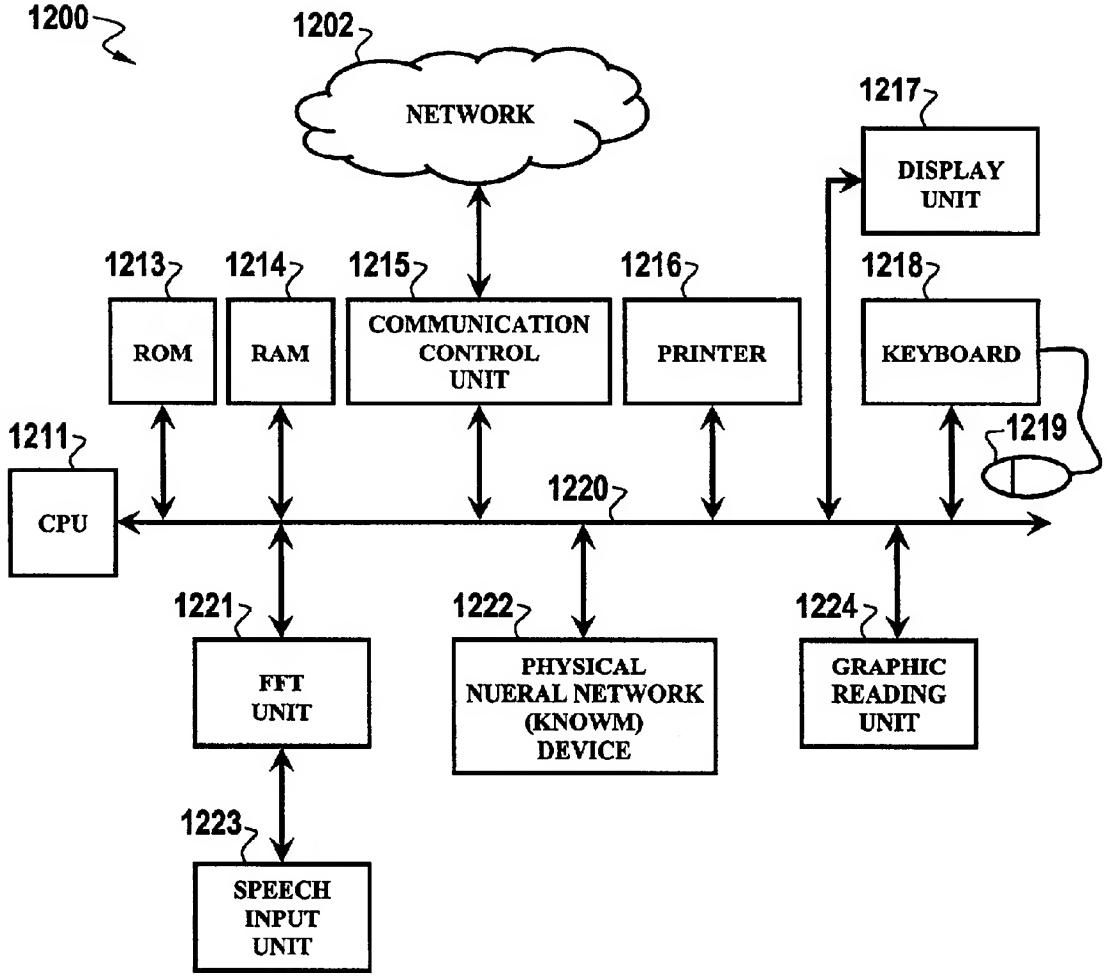


Figure 12

PHYSICAL NEURAL NETWORK DESIGN INCORPORATING NANOTECHNOLOGY

TECHNICAL FIELD

The present invention generally relates to nanotechnology. The present invention also relates to neural networks and neural computing systems and methods thereof. The present invention also relates to physical neural networks, which may be constructed based on nanotechnology. The present invention also related to VLSI (Very Large Scale Integrated) analog neural network chips. The present invention also relates to nanoconductors, such as nanotubes and nanowires.

BACKGROUND OF THE INVENTION

Neural networks are computational systems that permit computers to essentially function in a manner analogous to that of the human brain. Neural networks do not utilize the traditional digital model of manipulating 0's and 1's. Instead, neural networks create connections between processing elements, which are equivalent to neurons of a human brain. Neural networks are thus based on various electronic circuits that are modeled on human nerve cells (i.e., neurons). Generally, a neural network is an information-processing network, which is inspired by the manner in which a human brain performs a particular task or function of interest. Computational or artificial neural networks are thus inspired by biological neural systems. The elementary building block of biological neural systems is of course the neuron, the modifiable connections between the neurons, and the topology of the network.

Biologically inspired artificial neural networks have opened up new possibilities to apply computation to areas that were previously thought to be the exclusive domain of human intelligence. Neural networks learn and remember in ways that resemble human processes. Areas that show the greatest promise for neural networks, such as pattern classification tasks such as speech and image recognition, are areas where conventional computers and data-processing systems have had the greatest difficulty.

In general, artificial neural networks are systems composed of many nonlinear computational elements operating in parallel and arranged in patterns reminiscent of biological neural nets. The computational elements, or nodes, are connected via variable weights that are typically adapted during use to improve performance. Thus, in solving a problem, neural net models can explore many competing hypothesis simultaneously using massively parallel nets composed of many computational elements connected by links with variable weights. In contrast, with conventional von Neumann computers, an algorithm must first be developed manually, and a program of instructions written and executed sequentially. In some applications, this has proved extremely difficult. This makes conventional computers unsuitable for many real-time problems. A description and examples of artificial neural networks are disclosed in the publication entitled "Artificial Neural Networks Technology," by Dave Anderson and George McNeill, Aug. 10, 1992, a DACS (Data & Analysis Center for Software) State-of-the-Art Report under Contract Number F30602-89-C-0082, Rome Laboratory RL/C3C, Griffiss Air Force Base, New York.

In a neural network, "neuron-like" nodes can output a signal based on the sum of their inputs, the output being the result of an activation function. In a neural network, there exists a plurality of connections, which are electrically coupled among a plurality of neurons. The connections serve as communication bridges among of a plurality of neurons coupled thereto. A network of such neuron-like nodes has the ability to process information in a variety of useful ways. By adjusting the connection values between neurons in a network, one can match certain inputs with desired outputs.

One does not program a neural network. Instead, one "teaches" a neural network by examples. Of course, there are many variations. For instance, some networks do not require examples and extract information directly from the input data. The two variations are thus called supervised and unsupervised learning. Neural networks are currently used in applications such as noise filtering, face and voice recognition and pattern recognition. Neural networks can thus be utilized as an advanced mathematical technique for processing information.

Neural networks that have been developed to date are largely software-based. A true neural network (e.g., the human brain) is massively parallel (and therefore very fast computationally) and very adaptable. For example, half of a human brain can suffer a lesion early in its development and not seriously affect its performance. Software simulations are slow because during the learning phase a standard computer must serially calculate connection strengths. When the networks get larger (and therefore more powerful and useful), the computational time becomes enormous. For example, networks with 10,000 connections can easily overwhelm a computer. In comparison, the human brain has about 100 billion neurons, each of which is connected to about 5,000 other neurons. On the other hand, if a network is trained to perform a specific task, perhaps taking many days or months to train, the final useful result can be etched onto a piece of silicon and also mass-produced.

A number of software simulations of neural networks have been developed. Because software simulations are performed on conventional sequential computers, however, they do not take advantage of the inherent parallelism of neural network architectures. Consequently, they are relatively slow. One frequently used measurement of the speed of a neural network processor is the number of interconnections it can perform per second. For example, the fastest software simulations available can perform up to about 18 million interconnects per second. Such speeds, however, currently require expensive super computers to achieve. Even so, 18 million interconnects per second is still too slow to perform many classes of pattern classification tasks in real time. These include radar target classifications, sonar target classification, automatic speaker identification, automatic speech recognition and electro-cardiogram analysis, etc.

The implementation of neural network systems has lagged somewhat behind their theoretical potential due to the difficulties in building neural network hardware. This is primarily because of the large numbers of neurons and weighted connections required. The emulation of even of the simplest biological nervous systems would require neurons and connections numbering in the millions. Due to the difficulties in building such highly interconnected processors, the currently available neural network hardware systems have not approached this level of complexity. Another disadvantage of hardware systems is that they typically are often custom designed and built to implement one particular neural network architecture and are not easily, if at all, reconfigurable to implement different architectures. A true physical neural network chip, for example, has not yet been designed and successfully implemented.

The problem with pure hardware implementation of a neural network with technology as it exists today, is the

inability to physically form a great number of connections and neurons. On-chip learning can exist, but the size of the network would be limited by digital processing methods and associated electronic circuitry. One of the difficulties in creating true physical neural networks lies in the highly complex manner in which a physical neural network must be designed and built. The present inventor believes that solutions to creating a true physical and artificial neural network lies in the use of nanotechnology and the implementation of analog variable connections. The term "Nanotechnology" generally refers to nanometer-scale manufacturing processes, materials and devices, as associated with, for example, nanometer-scale lithography and nanometer-scale information storage. Nanometer-scale components find utility in a wide variety of fields, particularly in the fabrication of microelectrical and microelectromechanical systems (commonly referred to as "MEMS"). Microelectrical nano-sized components include transistors, resistors, capacitors and other nano-integrated circuit components. MEMS devices include, for example, micro-sensors, micro-actuators, micro-instruments, micro-optics, and the like.

In general, nanotechnology presents a solution to the problems faced in the rapid pace of computer chip design in recent years. According to Moore's law, the number of switches that can be produced on a computer chip has doubled every 18 months. Chips now can hold millions of transistors. However, it is becoming increasingly difficult to increase the number of elements on a chip using present technologies. At the present rate, in the next few years the theoretical limit of silicon based chips will be reached. Because the number of elements, which can be manufactured on a chip, determines the data storage and processing capabilities of microchips, new technologies are required which will allow for the development of higher performance chips.

Present chip technology is also limiting when wires need to be crossed on a chip. For the most part, the design of a computer chip is limited to two dimensions. Each time a circuit must cross another circuit, another layer must be added to the chip. This increases the cost and decreases the speed of the resulting chip. A number of alternatives to standard silicon based complementary metal oxide semiconductor ("CMOS") devices have been proposed. The common goal is to produce logic devices on a nanometer scale. Such dimensions are more commonly associated with molecules than integrated circuits.

Integrated circuits and electrical components thereof, which can be produced at a molecular and nanometer scale, include devices such as carbon nanotubes and nanowires, which essentially are nanoscale conductors ("nanoconductors"). Nanocondcutors are tiny conductive tubes (i.e., hollow) or wires (i.e., solid) with a very small size scale (e.g., 1.0–100 nanometers in diameter and hundreds of microns in length). Their structure and fabrication have been widely reported and are well known in the art. Carbon nanotubes, for example, exhibit a unique atomic arrangement, and possess useful physical properties such as one-dimensional electrical behavior, quantum conductance, and ballistic electron transport.

Carbon nanotubes are among the smallest dimensioned nanotube materials with a generally high aspect ratio and small diameter. High-quality single-walled carbon nanotubes can be grown as randomly oriented, needle-like or spaghetti-like tangled tubules. They can be grown by a number of fabrication methods, including chemical vapor deposition (CVD), laser ablation or electric arc growth. Carbon nanotubes can be grown on a substrate by catalytic decomposition of hydrocarbon containing precursors such as ethylene, methane, or benzene. Nucleation layers, such as thin coatings of Ni, Co, or Fe are often intentionally added onto the substrate surface in order to nucleate a multiplicity of isolated nanotubes. Carbon nanotubes can also be nucleated and grown on a substrate without a metal nucleating layer by using a precursor including one or more of these metal atoms. Semiconductor nanowires can be grown on substrates by similar processes.

Attempts have been made to construct electronic devices utilizing nano-sized electrical devices and components. For example, a molecular wire crossbar memory is disclosed in U.S. Pat. No. 6,128,214 entitled "Molecular Wire Crossbar Memory" dated Oct. 3, 2000 to Kuekes et al. Kuekes et al disclose a memory device that is constructed from crossbar arrays of nanowires sandwiching molecules that act as on/off switches. The device is formed from a plurality of nanometer-scale devices, each device comprising a junction formed by a pair of crossed wires where one wire crosses another and at least one connector species connects the pair of crossed wires in the junction. The connector species comprises a bi-stable molecular switch. The junction forms either a resistor or a diode or an asymmetric non-linear resistor. The junction has a state that is capable of being altered by application of a first voltage and sensed by the application of a second, non-destructive voltage. A series of related patents attempts to cover everything from molecular logic to how to chemically assemble these devices.

Such a molecular crossbar device has two general applications. The notion of transistors built from nanotubes and relying on nanotube properties is being pursued. Second, two wires can be selectively brought to a certain voltage and the resulting electrostatic force attracts them. When they touch, the Van der Walls force keeps them in contact with each other and a "bit" is stored. The connections in this apparatus can therefore be utilized for a standard (i.e., binary and serial) computer. The inventors of such a device thus desire to coax a nanoconductor into a binary storage media or a transistor. As it turns out, such a device is easier to utilize as a storage device.

The molecular wire crossbar memory device disclosed in Kuekes et al and related patents thereof simply comprise a digital storage medium that functions at a nano-sized level. Such a device, however, is not well-suited for non-linear and analog functions. Neural networks are non-linear in nature and naturally analog. A neural network is a very non-linear system, in that small changes to its input can create large changes in its output. To date, nanotechnology has not been applied to the creation of truly physical neural networks.

Based on the foregoing, the present inventor believes that a physical neural network which incorporates nanotechnology is a solution to the problems encountered by prior art neural network solutions. In particular, the present inventor believes that a true physical neural network can be designed and constructed without relying on computer simulations for training, or relying on standard digital (binary) memory to store connections strengths.

BRIEF SUMMARY OF THE INVENTION

The following summary of the invention is provided to facilitate an understanding of some of the innovative features unique to the present invention, and is not intended to be a full description. A full appreciation of the various aspects of the invention can be gained by taking the entire specification, claims, drawings, and abstract as a whole.

It is, therefore, one aspect of the present invention to provide a physical neural network.

It is therefore another aspect of the present to provide a physical neural network, which can be formed and implemented utilizing nanotechnology.

It is still another aspect of the present invention to provide a physical neural network, which can be formed from a plurality of interconnected nanoconnections or nanoconnectors.

It is a further aspect of the present invention to provide neuron like nodes, which can be formed and implemented utilizing nanotechnology;

It is also an aspect of the present invention to provide a physical neural network that can be formed from one or more neuron-like nodes.

It is yet a further aspect of the present invention to provide a physical neural network, which can be formed from a plurality of nanoconductors, such as, for example, nanowires and/or nanotubes.

It is still an additional aspect of the present invention to provide a physical neural network, which can be implemented physically in the form of a chip structure.

The above and other aspects can be achieved as is now described. A physical neural network based on nanotechnology is disclosed herein, including methods thereof. Such a physical neural network generally includes one or more neuron-like nodes, connected to a plurality of interconnected nanoconnections. Each neuron-like node sums one or more input signals and generates one or more output signals based on a threshold associated with the input signal. The physical neural network also includes a connection network formed from the interconnected nanoconnections, such that the interconnected nanoconnections used thereof by one or more of the neuron-like nodes can be strengthened or weakened according to an application of an electric field. Alignment has also been observed with a magnetic field, but electric fields are generally more practical. Note that the connection network is associated with one or more of the neuron-like nodes.

The output signal is generally based on a threshold below which the output signal is not generated and above which the output signal is generated. The transition from zero output to high output need not necessarily be abrupt or non linear. The connection network comprises a number of layers of nanoconnections, wherein the number of layers is equal to a number of desired outputs from the connection network. The nanoconnections are formed without influence from disturbances resulting from other nanoconnections thereof. Such nanoconnections may be formed from an electrically conducting material. The electrically conducting material is chosen such that a dipole is induced in the electrically conducting material in the presence of an electric field. Such a nanoconnection may comprise a nanoconductor.

The connection network itself may comprise a connection network structure having a connection gap formed therein, and a solution located within the connection gap, such that the solution comprises a solvent or suspension and one or more nanoconductors. Preferably, a plurality of nanoconductors is present in the solution (i.e., mixture). Note that such a solution may comprise a liquid and/or gas. An electric field can then be applied across the connection gap to permit the alignment of one or more of the nanoconductors within the connection gap. The nanoconductors can be suspended in the solvent, or can lie at the bottom of the connection gap on the surface of the chip. Studies have shown that nanotubes can align both in the suspension and/or on the surface of the gap. The electrical conductance of the mixture is less than the electrical conductance of the nanoconductors within the solution.

The nanoconductors within the connection gap thus experience an increased alignment in accordance with an increase in the electric field applied across the connection gap. Thus, nanoconnections of the neuron-like node that are utilized most frequently by the neuron-like node become stronger with each use thereof. The nanoconnections that are utilized least frequently become increasingly weak and eventually dissolve back into the solution. The nanoconnections may or may not comprise a resistance, which can be raised or lowered by a selective activation of a nanoconnection. They can be configured as nanoconductors such as, for example, a nanotube or nanowire. An example of a nanotube, which may be implemented in accordance with the invention described herein, is a carbon nanotube. Additionally, such nanoconnections may be configured as a negative connection associated with the neuron-like node.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying figures, in which like reference numerals refer to identical or functionally-similar elements throughout the separate views and which are incorporated in and form part of the specification, further illustrate the present invention and, together with the detailed description of the invention, serve to explain the principles of the present invention.

FIG. **1** illustrates a graph illustrating a typical activation function that can be implemented in accordance with the physical neural network of the present invention;

FIG. **2** depicts a schematic diagram illustrating a diode configuration as a neuron, in accordance with a preferred embodiment of the present invention;

FIG. **3** illustrates a block diagram illustrating a network of nanoconnections formed between two electrodes, in accordance with a preferred embodiment of the present invention;

FIG. **4** depicts a block diagram illustrating a plurality of connections between inputs and outputs of a physical neural network, in accordance with a preferred embodiment of the present invention;

FIG. **5** illustrates a schematic diagram of a physical neural network that can be created without disturbances, in accordance with a preferred embodiment of the present invention;

FIG. **6** depicts a schematic diagram illustrating an example of a physical neural network that can be implemented in accordance with an alternative embodiment of the present invention;

FIG. **7** illustrates a schematic diagram illustrating an example of a physical neural network that can be implemented in accordance with an alternative embodiment of the present invention;

FIG. **8** depicts a schematic diagram of a chip layout for a connection network that may be implemented in accordance with an alternative embodiment of the present invention;

FIG. **9** illustrates a flow chart of operations illustrating operational steps that may be followed to construct a connection network, in accordance with a preferred embodiment of the present invention;

FIG. **10** depicts a flow chart of operations illustrating operational steps that may be utilized to strengthen nanoconductors within a connection gap, in accordance with a preferred embodiment of the present invention;

FIG. **11** illustrates a schematic diagram of a circuit illustrating temporal summation within a neuron, in accordance with a preferred embodiment of the present invention; and

FIG. **12** depicts a block diagram illustrating a pattern recognition system, which may be implemented with a

physical neural network device, in accordance with an alternative embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

The particular values and configurations discussed in these non-limiting examples can be varied and are cited merely to illustrate an embodiment of the present invention and are not intended to limit the scope of the invention.

The physical neural network described and disclosed herein is different from prior art forms of neural networks in that the disclosed physical neural network does not require a computer simulation for training, nor is its architecture based on any current neural network hardware device. The design of the physical neural network of the present invention is actually quite "organic". The physical neural network described herein is generally fast and adaptable, no matter how large such a physical neural network becomes. The physical neural network described herein can be referred to generically as a Knowm. The terms "physical neural network" and "Knowm" can be utilized interchangeably to refer to the same device, network, or structure.

Network orders of magnitude larger than current VSLI neural networks can be built and trained with a standard computer. One consideration for a Knowm is that it must be large enough for its inherent parallelism to shine through. Because the connection strengths of such a physical neural network are dependant on the physical movement of nanoconnections thereof, the rate at which a small network can learn is generally very small and a comparable network simulation on a standard computer can be very fast. On the other hand, as the size of the network increases, the time to train the device does not change. Thus, even if the network takes a full second to change a connection value a small amount, if it does the same to a billion connections simultaneously, then its parallel nature begins to express itself.

A physical neural network (i.e., a Knowm) must have two components to function properly. First, the physical neural network must have one or more neuron-like nodes that sum a signal and output a signal based on the amount of input signal received. Such a neuron-like node is generally non-linear in its output. In other words, there should be a certain threshold for input signals, below which nothing is output and above which a constant or nearly constant output is generated or allowed to pass. This is a very basic requirement of standard software-based neural networks, and can be accomplished by an activation function. The second requirement of a physical neural network is the inclusion of a connection network composed of a plurality of interconnected connections (i.e., nanoconnections). Such a connection network is described in greater detail herein.

FIG. **1** illustrates a graph **100** illustrating a typical activation function that can be implemented in accordance with the physical neural network of the present invention. Note that the activation function need not be non-linear, although non-linearity is generally desired for learning complicated input-output relationships. The activation function depicted in FIG. **1** comprises a linear function, and is shown as such for general edification and illustrative purposes only. As explained previously, an activation function may also be non-linear.

As illustrated in FIG. **1**, graph **100** includes a horizontal axis **104** representing a sum of inputs, and a vertical axis **102** representing output values. A graphical line **106** indicates threshold values along a range of inputs from approximately −10 to +10 and a range of output values from approximately 0 to 1. As more neural networks (i.e., active inputs) are established, the overall output as indicated at line **105** climbs until the saturation level indicated by line **106** is attained. If a connection is not utilized, then the level of output (i.e., connection strength) begins to fade until it is revived. This phenomenon is analogous to short term memory loss of a human brain. Note that graph **100** is presented for generally illustrative and edification purposes only and is not considered a limiting feature of the present invention.

In a Knowm, the neuron-like node can be configured as a standard diode-based circuit, the diode being the most basic semiconductor electrical component, and the signal it sums may be a voltage. An example of such an arrangement of circuitry is illustrated in FIG. **2**, which generally depicts a schematic diagram illustrating a diode-based configuration as a neuron **200**, in accordance with a preferred embodiment of the present invention. Those skilled in the art can appreciate that the use of such a diode-based configuration is not considered a limiting feature of the present invention, but merely represents one potential arrangement in which the present invention may be implemented.

Although a diode may not necessarily be utilized, its current versus voltage characteristics are non-linear when used with associated resistors and similar to the relationship depicted in FIG. **1**. The use of a diode as a neuron is thus not a limiting feature of the present invention, but is only referenced herein with respect to a preferred embodiment. The use of a diode and associated resistors with respect to a preferred embodiment simply represents one potential "neuron" implementation. Such a configuration can be said to comprise an artificial neuron. It is anticipated that other devices and components may be utilized instead of a diode to construct a physical neural network and a neuron-like node (i.e., artificial neuron), as indicated here.

Thus, neuron **200** comprises a neuron-like node that may include a diode **206**, which is labeled D_1, and a resistor **204**, which is labeled R_2. Resistor **204** is connected to a ground **210** and an input **205** of diode **206**. Additionally, a resistor **202**, which is represented as a block and labeled R_1 can be connected to input **205** of diode **206**. Block **202** includes an input **212**, which comprises an input to neuron **200**. A resistor **208**, which is labeled R_3, is also connected to an output **214** of diode **206**. Additionally, resistor **208** is coupled to ground **210**. Diode **206** in a physical neural network is analogous to a neuron of a human brain, while an associated connection formed thereof, as explained in greater detail herein, is analogous to a synapse of a human brain.

As depicted in FIG. **2**, the output **214** is determined by the connection strength of R_1 (i.e., resistor **202**). If the strength of R_1's connection increases (i.e., the resistance decreases), then the output voltage at output **214** also increases. Because diode **206** conducts essentially no current until its threshold voltage (e.g., approximately 0.6V for silicon) is attained, the output voltage will remain at zero until R_1 conducts enough current to raise the pre-diode voltage to approximately 0.6V. After 0.6V has been achieved, the output voltage at output **214** will increase linearly. Simply adding extra diodes in series or utilizing different diode types may increase the threshold voltage.

An amplifier may also be added to the output **214** of diode **206** so that the output voltage immediately saturates at the diode threshold voltage, thus resembling a step function, until a threshold value and a constant value above the threshold is attained. R_3 (i.e., resistor **208**) functions gener-

ally as a bias for diode **206** (i.e., D_1) and should generally be about 10 times larger than resistor **204** (i.e., R_2). In the circuit configuration illustrated in FIG. **2**, R_1 can actually be configured as a network of connections composed of many inter-connected conducting nanowires (i.e., see FIG. **3**). As explained previously, such connections are analogous to the synapses of a human brain.

FIG. **3** illustrates a block diagram illustrating a network **300** of nanoconnections **304** formed between two electrodes, in accordance with a preferred embodiment of the present invention. Nanoconnections **304** (e.g., nanoconductors) depicted in FIG. **3** are generally located between input **302** and output **306**. The network of nanoconnections depicted in FIG. **3** can be implemented as a network of nanoconductors. Examples of nanoconductors include devices such as, for example, nanowires, nanotubes, and nanoparticles. Nanoconnections **304**, which are analogous to the synapses of a human brain, should be composed of electrical conducting material (i.e., nanoconductors). It should be appreciated by those skilled in the art that such nanoconductors can be provided in a variety of shapes and sizes without departing from the teachings herein.

For example, carbon particles (e.g., granules or bearings) may be used for developing nanoconnections. The nanoconductors utilized to form a connection network may be formed as a plurality of nanoparticles. For example, each nanoconnection within a connection network may be formed from a chain of carbon nanoparticles. In "Self-assembled chains of graphitized carbon nanoparticles" by Bezryadin et al., Applied Physics Letters, Vol. 74, No. 18, pp. 2699–2701, May 3, 1999, for example, a technique is reported, which permits the self-assembly of conducting nanoparticles into long continuous chains. Bezryadin et al. suggest that new approaches could be developed in order to organize nanoparticles into useful electronic devices. Thus, nanoconductors utilized to form a physical neural network (i.e., Knowm) could be formed from nanoparticles.

It should be appreciated by those skilled in the art that the Bezyadin et al reference does not, of course, comprise limiting features of the present invention, nor does it teach, suggest nor anticipate a physical neural network. Rather, such a reference merely demonstrate recent advances in the carbon nanotechnology arts and how such advances may be adapted for use in association with the Knowm-based system described herein. It can be further appreciated that a connection network as disclosed herein may be composed from a variety of different types of nanoconductors. For example, a connection network may be formed from a plurality of nanoconductors, including nanowires, nanotubes and/or nanoparticles. Note that such nanowires, nanotubes and/or nanoparticles, along with other types of nanoconductors can be formed from materials such as carbon or silicon. For example, carbon nanotubes may comprise a type of nanotube that can be utilized in accordance with the present invention.

As illustrated in FIG. **3**, nanoconnections **304** comprise a plurality of interconnected nanoconnections, which from this point forward, can be referred to generally as a "connection network." An individual nanoconnection may constitute a nanoconductor such as, for example, a nanowire, a nanotube, nanoparticles(s), or any other nanoconducting structures. Nanoconnections **304** may comprise a plurality of interconnected nanotubes and/or a plurality of interconnected nanowires. Similarly, nanoconnections **304** may be formed from a plurality of interconnected nanoparticles. A connection network is thus not one connection between two electrodes, but a plurality of connections between inputs and outputs. Nanotubes, nanowires, nanoparticles and/or other nanoconducting structures may be utilized, of course, to construct nanoconnections **304** between input **302** and input **306**. Although a single input **302** and a single input **306** is depicted in FIG. **3**, it can be appreciated that a plurality of inputs and a plurality of outputs may be implemented in accordance with the present invention, rather than simply a single input **302** or a single output **306**.

FIG. **4** depicts a block diagram illustrating a plurality of nanoconnections **414** between inputs **404**, **406**, **408**, **410**, **412** and outputs **416** and **418** of a physical neural network, in accordance with a preferred embodiment of the present invention. Inputs **404**, **406**, **408**, **410**, and **412** can provide input signals to connections **414**. Output signals can then be generated from connections **414** via outputs **416** and **418**. A connection network can therefore be configured from the plurality of connections **414**. Such a connection network is generally associated with one or more neuron-like nodes.

The connection network also comprises a plurality of interconnected nanoconnections, wherein each nanoconnection thereof is strengthened or weakened according to an application of an electric field. A connection network is not possible if built in one layer because the presence of one connection can alter the electric field so that other connections between adjacent electrodes could not be formed. Instead, such a connection network can be built in layers, so that each connection thereof can be formed without being influenced by field disturbances resulting from other connections. This can be seen in FIG. **5**.

FIG. **5** illustrates a schematic diagram of a physical neural network **500** that can be created without disturbances, in accordance with a preferred embodiment of the present invention. Physical neural network **500** is composed of a first layer **558** and a second layer **560**. A plurality of inputs **502**, **504**, **506**, **508**, and **510** are respectively provided to layers **558** and **560** respectively via a plurality of input lines **512**, **514**, **516**, **518**, and **520** and a plurality of input lines **522**, **524**, **526**, **528**, and **530**. Input lines **512**, **514**, **516**, **518**, and **520** are further coupled to input lines **532**, **534**, **536**, **538**, and **540** such that each line **532**, **534**, **536**, **538**, and **540** is respectively coupled to nanoconnections **572**, **574**, **576**, **578**, and **580**. Thus, input line **532** is connected to nanconnections **572**. Input line **534** is connected to nanoconnections **574**, and input line **536** is connected to nanoconnections **576**. Similarly, input line **538** is connected to nanconnections **578**, and input line **540** is connected to nanoconnections **580**.

Nanconnections **572**, **574**, **576**, **578**, and **580** may comprise nanoconductors such as, for example, nanotubes and/or nanowires. Nanoconnections **572**, **574**, **576**, **578**, and **580** thus comprise one or more nanoconductors. Additionally, input lines **522**, **524**, **526**, **528**, and **530** are respectively coupled to a plurality of input lines **542**, **544**, **546**, **548** and **550**, which are in turn each respectively coupled to nanoconnections **582**, **584**, **586**, **588**, and **590**. Thus, for example, input line **542** is connected to nanoconnections **582**, while input line **544** is connected to nanoconnections **584**. Similarly, input line **546** is connected to nanoconnections **586** and input line **548** is connected to nanoconnections **588**. Additionally, input line **550** is connected to nanconnections **590**. Box **556** and **554** generally represent simply the output and are thus illustrated connected to outputs **562** and **568**. In other words, outputs **556** and **554** respectively comprise outputs **562** and **568**. The aforementioned input lines and associated components thereof actually comprise physical electronic components, including conducting input and output lines and physical nanoconnections, such as nanotubes and/or nanowires.

Thus, the number of layers **558** and **560** equals the number of desired outputs **562** and **568** from physical neural network **500**. In the previous two figures, every input was potentially connected to every output, but many other configurations are possible. The connection network can be made of any electrically conducting material, although the physics of it requires that they be very small so that they will align with a practical voltage. Carbon nanotubes or any conductive nanowire can be implemented in accordance with the physical neural network described herein.

Such components can form connections between electrodes by the presence of an electric field. For example, the orientation and purification of carbon nanotubes has been demonstrated using ac electrophoresis in isopropyl alcohol, as indicated in "Orientation and purification of carbon nanotubes using ac electrophoresis" by Yamamoto et al., J. Phys. D: Applied Physics, 31(1998), L34–36. Additionally, an electric-field assisted assembly technique used to position individual nanowires suspended in an electric medium between two electrodes defined lithographically on an SiO2 substrate is indicated in "Electric-field assisted assembly and alignment of metallic nanowires," by Smith et al., Applied Physics Letters, Vol. 77, Num. 9, Aug. 28, 200. It can be appreciated by those skilled in the art that such references are not considered limiting features of the present invention, nor do such references teach, suggest or anticipate a physical neural network as described herein. Such references are discussed herein for general background and illustrative purposes only.

Additionally, it has been reported that it is possible to fabricate deterministic wiring networks from single-walled carbon nanotubes (SWNTs) as indicated in "Self-Assembled, Deterministic Carbon Nanotube Wiring Networks" by Diehi, et al. in Angew. Chem. Int. Ed. 2002, 41. No. 2. In addition, the publication "Indium phosphide nanowires as building blocks for nanoscale electronic and optoelectronic devices" by Duan, et al., Nature, Vol. 409, Jan. 4, 2001, reports that an electric-field-directed assembly can be used to create highly integrated device arrays from nanowire building blocks. It should be appreciated by those skilled in the art these references do not comprise limiting features of the present invention, nor do such references teach or anticipate a physical neural network. Rather, such references are incorporated herein by reference to demonstrate recent advances in the carbon nanotechnology arts and how such advances may be adapted for use in association with the physical neural network described herein.

The only general requirements for the conducting material utilized to configure the nanoconductors are that such conducting material must conduct electricity, and a dipole should preferably be induced in the material when in the presence of an electric field. Alternatively, the nanoconductors utilized in association with the physical neural network described herein can be configured to include a permanent dipole that is produced by a chemical means, rather than a dipole that is induced by an electric field.

Therefore, it should be appreciated by those skilled in the art that a connection network could also be comprised of other conductive particles that may be developed or found useful in the nanotechnology arts. For example, carbon particles (or "dust") may also be used as nanoconductors in place of nanowires or nanotubes. Such particles may include bearings or granule-like particles.

A connection network can be constructed as follows: A voltage is applied across a gap that is filled with a mixture of nanowires and a "solvent". This mixture could be made of many things. The only requirements are that the conducting wires must be suspended in the solvent, either dissolved or in some sort of suspension, free to move around; the electrical conductance of the substance must be less than the electrical conductance of the suspended conducting wire; and the viscosity of the substance should not be too much so that the conducting wire cannot move when an electric field is applied.

The goal for such a connection network is to develop a network of connections of just the right values so as to satisfy the particular signal-processing requirement—exactly what a neural network does. Such a connection network can be constructed by applying a voltage across a space occupied by the mixture mentioned. To create the connection network, the input terminals are selectively raised to a positive voltage while the output terminals are selectively grounded. Thus, connections can gradually form between the inputs and outputs. The important requirement that makes the physical neural network of the present invention functional as a neural network is that the longer this electric field is applied across a connection gap, or the greater the frequency or amplitude, the more nanotubes and/or nanowires and/or particles align and the stronger the connection thereof becomes. Thus, the connections that are utilized most frequently by the physical neural network become the strongest.

The connections can either be initially formed and have random resistances or no connections may be formed at all. By initially forming random connections, it might be possible to teach the desired relationships faster, because the base connections do not have to be built up from scratch. Depending on the rate of connection decay, having initial random connections could prove faster, although not necessarily. The connection network can adapt itself to the requirements of a given situation regardless of the initial state of the connections. Either initial condition will work, as connections that are not used will "dissolve" back into solution. The resistance of the connection can be maintained or lowered by selective activations of the connection. In other words, if the connection is not used, it will fade away, analogous to the connections between neurons in a human brain. The temperature of the solution can also be maintained at a particular value so that the rate that connections fade away can be controlled. Additionally an electric field can be applied perpendicular to the connections to weaken them, or even erase them out altogether (i.e., as in clear, zero, or reformatting of a "disk").

The nanoconnections may or may not be arranged in an orderly array pattern. The nanoconnections (e.g., nanotubes, nanowires, etc) of a physical neural network do not have to order themselves into neatly formed arrays. They simply float in the solution, or lie at the bottom of the gap, and more or less line up in the presence an electric field. Precise patterns are thus not necessary. In fact, neat and precise patterns may not be desired. Rather, due to the non-linear nature of neural networks, precise patterns could be a drawback rather than an advantage. In fact, it may be desirable that the connections themselves function as poor conductors, so that variable connections are formed thereof, overcoming simply an "on" and "off" structure, which is commonly associated with binary and serial networks and structures thereof.

FIG. **6** depicts a schematic diagram illustrating an example of a physical neural network **600** that can be implemented in accordance an alternative embodiment of the present invention. Note that in FIGS. **5** and **6**, like parts are indicated by like reference numerals. Thus, physical

neural network **600** can be configured, based on physical neural network **500** illustrated in FIG. **5**. In FIG. **6**, inputs **1**, **2**, **3**, **4**, and **5** are indicated, which are respectively analogous to inputs **502**, **504**, **506**, **508**, and **510** illustrated in FIG. **5**. Outputs **562** and **568** are provided to a plurality of electrical components to create a first output **626** (i.e., Output **1**) and a second output **628** (i.e., Output **2**). Output **562** is tied to a resistor **606**, which is labeled R2 and a diode **616** at node A. Output **568** is tied to a resistor **610**, which is also labeled R2 and a diode **614** at node C. Resistors **606** and **610** are each tied to a ground **602**.

Diode **616** is further coupled to a resistor **608**, which is labeled R3, and first output **626**. Additionally, resistor **608** is coupled to ground **602** and an input to an amplifier **618**. An output from amplifier **618**, as indicated at node B and dashed lines thereof, can be tied back to node A. A desired output **622** from amplifier **618** is coupled to amplifier **618** at node H. Diode **614** is coupled to a resistor **612** at node F. Note that resistor **612** is labeled R3. Node F is in turn coupled to an input of amplifier **620** and to second output **628** (i.e., Output **2**). Diode **614** is also connected to second output **628** and an input to amplifier **620** at second output **628**. Note that second output **628** is connected to the input to amplifier **620** at node F. An output from amplifier **620** is further coupled to node D, which in turn is connected to node C. A desired output **624**, which is indicated by a dashed line in FIG. **6**, is also coupled to an input of amplifier **620** at node E.

In FIG. **6**, the training of physical neural network **600** can be accomplished utilizing, for example, op-amp devices (e.g., amplifiers **618** and **620**). By comparing an output (e.g., first output **626**) of physical neural network **600** with a desired output (e.g., desired output **622**), the amplifier (e.g., amplifier **618**) can provide feedback and selectively strengthen connections thereof. For instance, suppose it is desired to output a voltage of +V at first output **626** (i.e., Output **1**) when inputs **1** and **4** are high. When inputs **1** and **4** are taken high, also assume that first output **626** is zero. Amplifier **618** can then compare the desired output (+V) with the actual output (0) and output −V. In this case, −V is equivalent to ground.

The op-amp outputs and grounds the pre-diode junction (i.e., see node A) and causes a greater electric field across inputs **1** and **4** and the layer **1** output. This increased electric field (larger voltage drop) can cause the nanoconductors in the solution between the electrode junctions to align themselves, aggregate, and form a stronger connection between the **1** and **4** electrodes. Feedback can continue to be applied until output of physical neural network **600** matches the desired output. The same procedure can be applied to every output.

In accordance with the aforementioned example, assume that Output **1** was higher than the desired output (i.e., desired output **622**). If this were the case, the op-amp output can be +V and the connection between inputs **1** and **4** and layer one output can be raised to +V. Columbic repulsions between the nanoconductors can force the connection apart, thereby weakening the connection. The feedback will then continue until the desired output is obtained. This is just one training mechanism. One can see that the training mechanism does not require any computations, because it is a simple feedback mechanism.

Such a training mechanism, however, may be implemented in many different forms. Basically, the connections in a connection network must be able to change in accordance with the feedback provided. In other words, the very general notion of connections being strengthened or connections being weakened in a physical system is the essence of a physical neural network (i.e., Knowm). Thus, it can be appreciated that the training of such a physical neural network may not require a "CPU" to calculate connection values thereof. The Knowm can adapt itself. Complicated neural network solutions could be implemented very rapidly "on the fly", much like a human brain adapts as it performs.

The physical neural network disclosed herein thus has a number of broad applications. The core concept of a Knowm, however, is basic. The very basic idea that the connection values between electrode junctions by nanoconductors can be used in a neural network devise is all that required to develop an enormous number of possible configurations and applications thereof.

Another important feature of a physical neural network is the ability to form negative connections. This is an important feature that makes possible inhibitory effects useful in data processing. The basic idea is that the presence of one input can inhibit the effect of another input. In artificial neural networks as they currently exist, this is accomplished by multiplying the input by a negative connection value. Unfortunately, with a Knowm-based device, the connection may only take on zero or positive values under such a scenario.

In other words, either there can be a connection or no connection. A connection can simulate a negative connection by dedicating a particular connection to be negative, but one connection cannot begin positive and through a learning process change to a negative connection. In general, if it starts positive, it can only go to zero. In essence, it is the idea of possessing a negative connection initially that results in the simulation, because this does not occur in a human brain. Only one type of signal generally travels through axon/dendrites in a human brain. That signal is transferred into the flow of a neurotransmitter whose effect on the postsynaptic neuron can be either excitatory or inhibitory, depending on the neuron.

One method for solving this problem is to utilize two sets of connections for the same output, having one set represent the positive connections and the other set represent the negative connections. The output of these two layers can be compared, and the layer with the greater output will output either a high signal or a low signal, depending on the type of connection set (inhibitory or excitatory). This can be seen in FIG. **7**.

FIG. **7** illustrates a schematic diagram illustrating an example of a physical neural network **700** that can be implemented in accordance with an alternative embodiment of the present invention. Physical neural network **700** thus comprises a plurality of inputs **702** (not necessarily binary) which are respectively fed to layers **704**, **706**, **708**, and **710**. Each layer is analogous to the layers depicted earlier, such as for example layers **558** and **560** of FIG. **5**. An output **713** of layer **704** can be connected to a resistor **712**, a transistor **720** and a first input **727** of amplifier **726**. Transistor **720** is generally coupled between ground **701** and first input **727** of amplifier **726**. Resistor **712** is connected to a ground **701**. Note that ground **701** is analogous to ground **602** illustrated in FIG. **6** and ground **210** depicted in FIG. **2**. A second input **729** of amplifier **726** can be connected to a threshold voltage 756. The output of amplifier **726** can in turn be fed to an inverting amplifier **736**.

The output of inverting amplifier **736** can then be input to a NOR device **740**. Similarly, an output **716** of layer **706** may be connected to resistor **714**, transistor **733** and a first input **733** of an amplifier **728**. A threshold voltage 760 is

connected to a second input **737** of amplifier **728**. Resistor **714** is generally coupled between ground **701** and first input **733** of amplifier **728**. Note that first input **733** of amplifier **728** is also generally connected to an output **715** of layer **706**. The output of amplifier **728** can in turn be provided to NOR device **740**. The output from NOR device **740** is generally connected to a first input **745** of an amplifier **744**. An actual output **750** can be taken from first input **745** to amplifier **744**. A desired output **748** can be taken from a second input **747** to amplifier **744**. The output from amplifier **744** is generally provided at node A, which in turn is connected to the input to transistor **720** and the input to transistor **724**. Note that transistor **724** is generally coupled between ground **701** and first input **733** of amplifier **728**. The second input **731** of amplifier **728** can produce a threshold voltage 760.

Layer **708** provides an output **717** that can be connected to resistor **716**, transistor **725** and a first input **737** to an amplifier **732**. Resistor **716** is generally coupled between ground **701** and the output **717** of layer **708**. The first input **737** of amplifier **732** is also electrically connected to the output **717** of layer **708**. A second input **735** to amplifier **732** may be tied to a threshold voltage 758. The output from amplifier **732** can in turn be fed to an inverting amplifier **738**. The output from inverting amplifier **738** may in turn be provided to a NOR device **742**. Similarly, an output **718** from layer **710** can be connected to a resistor **719**, a transistor **728** and a first input **739** of an amplifier **734**. Note that resistor **719** is generally coupled between node **701** and the output **719** of layer **710**. A second input **741** of amplifier **734** may be coupled to a threshold voltage 762. The output from of NOR device **742** is generally connected to a first input **749** of an amplifier **746**. A desired output **752** can be taken from a second input **751** of amplifier **746**. An actual output **754** can be taken from first input **749** of amplifier **746**. The output of amplifier **746** may be provided at node B, which in turn can be tied back to the respective inputs to transistors **725** and **728**. Note that transistor **725** is generally coupled between ground **701** and the first input **737** of amplifier **732**. Similarly, transistor **728** is generally connected between ground **701** and the first input **739** of amplifier **734**.

Note that transistors **720**, **724**, **725** and/or **728** each can essentially function as a switch to ground. A transistor such as, for example, transistor **720**, **724**, **725** and/or **728** may comprise a field-effect transistor (FET) or another type of transistor, such as, for example, a single-electron transistor (SET). Single-electron transistor (SET) circuits are essential for hybrid circuits combining quantum SET devices with conventional electronic devices. Thus, SET devices and circuits may be adapted for use with the physical neural network of the present invention. This is particularly important because as circuit design rules begin to move into regions of the sub-**100** nanometer scale, where circuit paths are only 0.001 of the thickness of a human hair, prior art device technologies will begin to fail, and current leakage in traditional transistors will become a problem. SET offers a solution at the quantum level, through the precise control of a small number of individual electrons.

Transistors such as transistors **720**, **724**, **725** and/or **728** can also be implemented as carbon nanotube transistors. An example of a carbon nanotube transistor is disclosed in U.S. Patent Application No. 2001/0023986A1 to Macevski, which is dated Sep. 27, 2001 and is entitled, "System and Method for Fabricating Logic Devices Comprising Carbon Nanotube Transistors." U.S. Patent Application No. 2001/0023986A1 does not teach or claim a physical neural network, but instead teaches the formation of a discrete carbon nanotube transistor. Thus, U.S. Patent Application No. 2001/0023986A1 is not considered a limiting feature of the present invention, nor does this reference teach, anticipate or suggest the invention described herein. Instead, this reference is discussed briefly herein for background purposes only and to generally illustrate the use of a particular type of discrete transistor in the nanodomain.

A truth table for the output of circuit **700** is illustrated at block **780** in FIG. 7. As indicated at block **780**, when an excitatory output is high and the inhibitory output is also high, the final output is low. When the excitatory output is high and the inhibitory output is low, the final output is high. Similarly, when the excitatory output is low and the inhibitory output is high, the final output is low. When the excitatory output is low and the inhibitory output is also low, the final output is low. Note that layers **704** and **708** may thus comprise excitatory connections, while layers **706** and **710** may comprise inhibitory connections.

For every desired output, two sets of connections are used. The output of a two-diode neuron can be fed into an op-amp (e.g., a comparator). If the output that the op-amp receives is low when it should be high, the op-amp outputs a low signal. This low signal can cause the transistors (e.g., transistors **720**, **725**) to saturate and ground out the pre-diode junction for the excitatory diode. This causes, like before, an increase in the voltage drop across those connections that need to increase their strength. Note that only those connections going to the excitatory diode are strengthened. Likewise, if the desired output were low when the actual output was high, the op-amp can output a high signal. This can cause the inhibitory transistor (e.g., an NPN transistor) to saturate and ground out the neuron junction of the inhibitory connections. Connections going to the inhibitory diode can thereafter strengthen.

At all times during the learning process, a weak alternating electric field can be applied perpendicular to the connections. This can cause the connections to weaken by rotating the nanotube perpendicular to the connection direction. This perpendicular field is important because it can allow for a much higher degree of adaptation. To understand this, one must realize that the connections cannot (practically) keep getting stronger and stronger. By weakening those connections not contributing much to the desired output, we decrease the necessary strength of the needed connections and allow for more flexibility in continuous training. This perpendicular alternating voltage can be realized by the addition of two electrodes on the outer extremity of the connection set, such as plates sandwiching the connections (i.e., above and below). Other mechanisms, such as increasing the temperature of the nanotube suspension could also be used for such a purpose, although this method is perhaps a little less controllable or practical.

The circuit depicted in FIG. 7 can be separated into two separate circuits. The first part of the circuit can be composed of nanotube connections, while the second part of the circuit comprises the "neurons" and the learning mechanism (i.e., op-amps/comparator). The learning mechanism on first glance appears similar to a relatively standard circuit that could be implemented on silicon with current technology. Such a silicon implementation can thus comprise the "neuron" chip. The second part of the circuit (i.e., the connections) is thus a new type of chip, although it could be constructed with current technology. The connection chip can be composed of an orderly array of electrodes spaced anywhere from, for example, 100 nm to 1 μm or perhaps even further. In a biological system, one talks of synapses

connecting neurons. It is in the synapses where the information is processed, (i.e., the "connection weights"). Similarly, such a chip can contain all of the synapses for the physical neural network. A possible arrangement thereof can be seen in FIG. **8**.

FIG. **8** thus illustrates a possible chip layout for a connection chip (i.e., connection network **800**) that can be implemented in accordance with the present invention. FIG. **8** thus illustrates a possible chip layout for a connection chip (i.e., s connection network **800** that can be implemented in accordance with the present invention. Chip layout **800** can include an input array composed of a plurality of inputs **801**, **802**, **803**, **804**, and **805**, which are generally provided to a plurality of layers **806**, **807**, **808**, **809**, **810**, **811**, **812**, **813**, **814**, and **815**. A plurality of outputs **802** can be derived from layers **806**, **807**, **808**, **809**, **810**, **811**, **812**, **813**, **814**, and **815**. Inputs **801** can be coupled to layers **806** and **807**, while inputs **802** can be connected to layers **808** and **809**. Similarly, inputs **803** can be connected to layers **810** and **811**. Also, inputs **804** are generally connected to layers **812** and **813**. Inputs **805** are generally connected to layers **814** and **815**.

Similarly, such an input array can includes a plurality of inputs **831**, **832**, **833**, **834** and **835** which are respectively input to a plurality of layers **816**, **817**, **818**, **819**, **820**, **821**, **822**, **823**, **824** and **825**. Thus, inputs **831** are connected to layers **816** and **817**, while inputs **832** are coupled to layers **818** and **819**. Additionally, inputs **833** are connected to layers **820** and **821**. Inputs **834** are connected to layers **822** and **823**. Finally, inputs **835** are connected to layers **824** and **825**. Arrows **828** and **830** represent a continuation of the aforementioned connection network pattern. Those skilled in the art can appreciate, of course, that chip layout **800** is not intended to represent an exhaustive chip layout or to limit the scope of the invention. Many modifications and variations to chip layout **800** are possible in light of the teachings herein without departing from the scope of the present invention. It is contemplated that the use of a chip layout, such as chip layout **800**, can involve a variety of components having different characteristics.

Preliminary calculations by the present inventor based on a maximum etching capability of 200 nm resolution have indicated that over 4 million synapses could fit on an area of approximately 1 cm^2. The smallest width that an electrode can possess is generally based on current lithography. Such a width may of course change as the lithographic arts advance. This value is actually about 70 nm for state-of-the-art techniques currently. These calculations are of course extremely conservative, and are not considered a limiting feature of the present invention. Such calculations are based on an electrode with, separation, and gap of approximately 200 nm. For such a calculation, 166 connection networks comprising 250 inputs and 100 outputs can fit within a one square centimeter area.

If such chips are stacked vertically, an untold number of synapses could be attained. This is two to three orders of magnitude greater than some of the most capable neural network chips out there today, chips that rely on standard methods to calculate synapse weights. Of course, the geometry of the chip could take on many different forms, and it is quite possible (i.e., based on a conservative lithography and chip layout) that many more synapses could fit within the same space. The training of a neural network chip of this size would take a fraction of the time of a comparably sized traditional chip using digital technology.

The training of such a chip is primarily based on two assumptions. First, the inherent parallelism of a physical neural network (i.e., a Knowm) can permit all training sessions to occur simultaneously, no matter now large the associated connection network. Second, recent research has indicated that near perfect aligning of nanotubes can be accomplished, for example, in approximately 15 minutes. If one considers that the input data, arranged as a vector of binary "highs and low's" is presented to the Knowm simultaneously, and that all training vectors are presented one after the other in rapid succession (e.g., 100 MHz or more), then each connection would "see" a different frequency in direct proportion to the amount of time that its connection is required for accurate data processing (i.e., provided by a feedback mechanism). Thus, if it only takes for example, approximately 15 minutes to attain an almost perfect state of alignment, then this amount of time would comprise the longest amount of time required to train, assuming that all of the training vectors are presented during that particular time period.

FIG. **9** illustrates a flow chart **900** of operations illustrating operational steps that may be followed to construct a connection network, in accordance with a preferred embodiment of the present invention. Initially, as indicated at block **902**, a connection gap is created from a connection network structures. As indicated earlier, the goal for such a connection network is generally to develop a network of connections of "just" the right values to satisfy particular information processing requirements, which is precisely what a neural network accomplishes. As illustrated at block **904**, a solution is prepared, which is composed of nanoconductors and a "solvent." Note that the term "solvent" as utilized herein has a variable meaning, which includes the traditional meaning of a "solvent," and also a suspension.

The solvent utilized can comprise a volatile liquid that can be confined or sealed and not exposed to air. For example, the solvent and the nanoconductors present within the resulting solution may be sandwiched between wafers of silicon or other materials. If the fluid has a melting point that is approximately at room temperature, then the viscosity of the fluid could be controlled easily. Thus, if it is desired to lock the connection values into a particular state, the associated physical neural network (i.e., Knowm) may be cooled slightly until the fluid freezes. The term "solvent" as utilized herein thus can include fluids such as for example, toluene, hexadecane, mineral oil, etc. Note that the solution in which the nanoconductors (i.e., nanoconnections) are present should generally comprise a dielectric. Thus, when the resistance between the electrodes is measured, the conductivity of the nanoconductors is essentially measured, not that of the solvent. The nanoconductors can be suspended in the solution or can alternately lie on the bottom surface of the connection gap. The solvent may also be provided in the form of a gas.

As illustrated thereafter at block **906**, the nanoconductors must be suspended in the solvent, either dissolved or in a suspension of sorts, but generally free to move around, either in the solution or on the bottom surface of the gap. As depicted next at block **908**, the electrical conductance of the solution must be less than the electrical conductance of the suspended nanoconductor(s). Similarly, the electrical resistance of the solution is greater than the electrical resistance of the nanoconductor.

Next, as illustrated at block **910**, the viscosity of the substance should not be too much so that the nanoconductors cannot move when an electric field (e.g., voltage) is applied. Finally, as depicted at block **912**, the resulting solution of the "solvent" and the nanoconductors is thus located within the connection gap.

Note that although a logical series of steps is illustrated in FIG. **9**, it can be appreciated that the particular flow of steps can be re-arranged. Thus, for example, the creation of the connection gap, as illustrated at block **902**, may occur after the preparation of the solution of the solvent and nanoconductor(s), as indicated at block **904**. FIG. **9** thus represents merely possible series of steps, which may be followed to create a connection network. It is anticipated that a variety of other steps may be followed as long as the goal of achieving a connection network in accordance with the present invention is achieved. Similar reasoning also applies to FIG. **10**.

FIG. **10** depicts a flow chart **1000** of operations illustrating operational steps that may be utilized to strengthen nanoconductors within a connection gap, in accordance with a preferred embodiment of the present invention. As indicated at block **1002**, an electric field can be applied across the connection gap discussed above with respect to FIG. **9**. The connection gap can be occupied by the solution discussed above. As indicated thereafter at block **1004**, to create the connection network, the input terminals can be selectively raised to a positive voltage while the output terminals are selectively grounded. As illustrated thereafter at block **1006**, connections thus form between the inputs and the outputs. The important requirements that make the resulting physical neural network functional as a neural network is that the longer this electric field is applied across the connection gap, or the greater the frequency or amplitude, the more nanoconductors align and the stronger the connection becomes. Thus, the connections that get utilized the most frequently become the strongest.

As indicated at block **1008**, the connections can either be initially formed and have random resistances or no connections will be formed at all. By forming initial random connections, it might be possible to teach the desired relationships faster, because the base connections do not have to be built up as much. Depending on the rate of connection decay, having initial random connections could prove to be a faster method, although not necessarily. A connection network will adapt itself to whatever is required regardless of the initial state of the connections. Thus, as indicated at block **1010**, as the electric field is applied across the connection gap, the more the nonconductor(s) will align and the stronger the connection becomes. Connections (i.e., synapses) that are not used are dissolved back into the solution, as illustrated at block **1012**. As illustrated at block **1014**, the resistance of the connection can be maintained or lowered by selective activations of the connections. In other words, "if you do not use the connection, it will fade away," much like the connections between neurons in a human brain.

The neurons in a human brain, although seemingly simple when viewed individually, interact in a complicated network that computes with both space and time. The most basic picture of a neuron, which is usually implemented in technology, is a summing device that adds up a signal. Actually, this statement can be made even more general by stating that a neuron adds up a signal in discrete units of time. In other words, every group of signals incident upon the neuron can be viewed as occurring in one moment in time. Summation thus occurs in a spatial manner. The only difference between one signal and another signal depends on where such signals originate. Unfortunately, this type of data processing excludes a large range of dynamic, varying situations that cannot necessarily be broken up into discrete units of time.

The example of speech recognition is a case in point. Speech occurs in the time domain. A word is understood as the temporal pronunciation of various syllables. A sentence is composed of the temporal separation of varying words. Thoughts are composed of the temporal separation of varying sentences. Thus, for an individual to understand a spoken language at all, a syllable, word, sentence or thought must exert some type of influence on another syllable, word, sentence or thought. The most natural way that one sentence can exert any influence on another sentence, in the light of neural networks, is by a form of temporal summation. That is, a neuron "remembers" the signals it received in the past.

The human brain accomplishes this feat in an almost trivial manner. When a signal reaches a neuron, the neuron has an influx of ions rush through its membrane. The influx of ions contributes to an overall increase in the electrical potential of the neuron. Activation is achieved when the potential inside the cell reaches a certain threshold. The one caveat is that it takes time for the cell to pump out the ions, something that it does at a more or less constant rate. So, if another signal arrives before the neuron has time to pump out all of the ions, the second signal will add with the remnants of the first signal and achieve a raised potential greater than that which could have occurred with only the second signal. The first signal influences the second signal, which results in temporal summation.

Implementing this in a technological manner has proved difficult in the past. Any simulation would have to include a "memory" for the neuron. In a digital representation, this requires data to be stored for every neuron, and this memory would have to be accessed continually. In a computer simulation, one must discritize the incoming data, since operations (such as summations and learning) occur serially. That is, a computer can only do one thing at a time. Transformations of a signal from the time domain into the spatial domain require that time be broken up into discrete lengths, something that is not necessarily possible with real-time analog signals in which no point exists within a time-varying signal that is uninfluenced by another point.

A physical neural network, however, is generally not digital. A physical neural network is a massively parallel analog device. The fact that actual molecules (e.g., nanoconductors) must move around (in time) makes temporal summation a natural occurrence. This temporal summation is built into the nanoconnections. The easiest way to understand this is to view the multiplicity of nanoconnections as one connection with one input into a neuron-like node (Op-amp, Comparator, etc.). This can be seen in FIG. **11**.

FIG. **11** illustrates a schematic diagram of a circuit **1100** illustrating temporal summation within a neuron, in accordance with a preferred embodiment of the present invention. As indicated in FIG. **11**, an input **1102** can be provided to a nanoconnections **1104**, which in turn can provide a signal, which can be input to an amplifier **1110** (e.g., op amp) at node B. A resistor **1106** is connected to node A, which in turn is electrically equivalent to node B. Node B is connected to a negative input of amplifier **1100**. Resistor **1108** is also connected to a ground **1108**. Amplifier **1110** provides output **1114**. Note that although nanoconnections **1104** is referred to in the plural it can be appreciated that nanoconnections **1104** can comprise a single nanoconnection or a plurality of nanoconnections. For simplicity sake, however, the plural form is used to refer to nanoconnections **1104**.

Input **1102** can be provided by another physical neural network (i.e., Knowm) to cause increased connection strength of nanoconnections **1104** over time. This input would most likely arrive in pulses, but could also be

continuous. A constant or pulsed electric field perpendicular to the connections would serve to constantly erode the connections, so that only signals of a desired length or amplitude could cause a connection to form. Once the connection is formed, the voltage divider formed by nanoconnection **1104** and resistor **1106** can cause a voltage at node A in direct proportion to the strength of nanoconnections **1104**. When the voltage at node A reaches a desired threshold, the amplifier (i.e., an op-amp and/or comparator), will output a high voltage (i.e., output **1114**). The key to the temporal summation is that, just like a real neuron, it takes time for the electric field to breakdown the nanoconnections **1104**, so that signals arriving close in time will contribute to the firing of the neuron (i.e., op-amp, comparator, etc.). Temporal summation has thus been achieved. The parameters of the temporal summation could be adjusted by the amplitude and frequency of the input signals and the perpendicular electric field.

FIG. **12** depicts a block diagram illustrating a pattern recognition system **1200**, which may be implemented with a physical neural network device **1222**, in accordance with an alternative embodiment of the present invention. Note that pattern recognition system **1200** can be implemented as a speech recognition system. Those skilled in the art can appreciate, however, that although pattern recognition system **1200** is depicted herein in the context of speech recognition, a physical neural network device (i.e., a Knowm device) may be implemented with other pattern recognition systems, such as visual and/or imaging recognition systems. FIG. **12** thus does not comprise a limiting feature of the present invention and is presented for general edification and illustrative purposes only. Those skilled in the art can appreciate that the diagram depicted in FIG. **12** may be modified as new applications and hardware are developed. The development or use of a pattern recognition system such as pattern recognition system **1200** of FIG. **12** by no means limits the scope of the physical neural network (i.e., Knowm) disclosed herein.

FIG. **12** thus illustrates in block diagram fashion, the system structure of a speech recognition device using a neural network according to an alternative embodiment of the present invention. The pattern recognition system **1200** is provided with a CPU **1211** for performing the functions of inputting vector rows and instructor signals (vector rows) to an output layer for the learning process of a physical neural network device **1222**, and changing connection weights between respective neuron devices based on the learning process. Pattern recognition system **1200** can be implemented within the context of a data-processing system, such as, for example, a personal computer or personal digital assistant (PDA), both of which are well known in the art.

The CPU **1211** can perform various processing and controlling functions, such as pattern recognition, including but not limited to speech and/or visual recognition based on the output signals from the physical neural network device **1222**. The CPU **1211** is connected to a read-only memory (ROM) **1213**, a random-access memory (RAM) **1214**, a communication control unit **1215**, a printer **1216**, a display unit **1217**, a keyboard **1218**, an FFT (fast Fourier transform) unit **1221**, a physical neural network device **1222** and a graphic reading unit **1224** through a bus line **1220** such as a data bus line. The bus line **1220** may comprise, for example, an ISA, EISA, or PCI bus.

The ROM **1213** is a read-only memory storing various programs or data used by the CPU **1211** for performing processing or controlling the learning process, and speech recognition of the physical neural network device **1222**. The ROM **1213** may store programs for carrying out the learning process according to error back-propagation for the physical neural network device or code rows concerning, for example, 80 kinds of phonemes for performing speech recognition. The code rows concerning the phonemes can be utilized as second instructor signals and for recognizing phonemes from output signals of the neuron device network. Also, the ROM **1213** can store programs of a transformation system for recognizing speech from recognized phonemes and transforming the recognized speech into a writing (i.e., written form) represented by characters.

A predetermined program stored in the ROM **1213** can be downloaded and stored in the RAM **1214**. RAM **1214** generally functions as a random access memory used as a working memory of the CPU **1211**. In the RAM **1214**, a vector row storing area can be provided for temporarily storing a power obtained at each point in time for each frequency of the speech signal analyzed by the FFT unit **1221**. A value of the power for each frequency serves as a vector row input to a first input portion of the physical neural network device **1222**. Further, in the case where characters or graphics are recognized in the physical neural network device, the image data read by the graphic reading unit **1224** are stored in the RAM **1214**.

The communication control unit **1215** transmits and/or receives various data such as recognized speech data to and/or from another communication control unit through a communication network **1202** such as a telephone line network, an ISDN line, a LAN, or a personal computer communication network. Network **1202** may also comprise, for example, a telecommunications network, such as a wireless communications network. Communication hardware methods and systems thereof are well known in the art.

The printer **1216** can be provided with a laser printer, a bubble-type printer, a dot matrix printer, or the like, and prints contents of input data or the recognized speech. The display unit **1217** includes an image display portion such as a CRT display or a liquid crystal display, and a display control portion. The display unit **1217** can display the contents of the input data or the recognized speech as well as a direction of an operation required for speech recognition utilizing a graphical user interface (GUI).

The keyboard **1218** generally functions as an input unit for varying operating parameters or inputting setting conditions of the FFT unit **1221**, or for inputting sentences. The keyboard **1218** is generally provided with a ten-key numeric pad for inputting numerical figures, character keys for inputting characters, and function keys for performing various functions. A mouse **1219** can be connected to the keyboard **1218** and serves as a pointing device.

A speech input unit **1223**, such as a microphone can be connected to the FFT unit **1221**. The FFT unit **1221** transforms analog speech data input from the voice input unit **1223** into digital data and carries out spectral analysis of the digital data by discrete Fourier transformation. By performing a spectral analysis using the FFT unit **1221**, the vector row based on the powers of the respective frequencies are output at predetermined intervals of time. The FFT unit **1221** performs an analysis of time-series vector rows, which represent characteristics of the inputted speech. The vector rows output by the EFT **1221** are stored in the vector row storing area in the RAM **1214**. The graphic reading unit **224**, provided with devices such as a CCD (Charged Coupled Device), can be used for reading images such as characters or graphics recorded on paper or the like. The image data read by the image-reading unit **1224** are stored in the RAM

1214. Note that an example of a pattern recognition apparatus, which may be modified for use with the physical neural network of the present invention, is disclosed in U.S. Pat. No. 6,026,358 to Tomabechi, Feb. 16, 2000, "Neural Network, A Method of Learning of a Neural Network and Phoneme Recognition Apparatus Utilizing a Neural Network." It can be appreciated by those skilled in the art that the Tomabechi reference does not teach, suggest or anticipate the invention disclosed herein. The Tomabechi reference is discussed herein for illustrative, background general edification purposes only and is not considered a limiting feature of the present invention.

The implications of a physical neural network are tremendous. With existing lithography technology, many electrodes in an array such as depicted in FIG. **5** can be etched onto a wafer of silicon. The neurons (i.e., op-amps, diodes, etc.), as well as the training circuitry illustrated in FIG. **6**, could be built onto the same silicon wafer, although it may be desirable to have the connections on a separate chip due to the liquid solution of nanoconductors. A solution of suspended nanoconductors could be placed between the electrode connections and the chip could be packaged. The resulting "chip" would look much like a current Integrated Chip (IC) or VLSI (very large scale integrated) chips. One could also place a rather large network parallel with a computer processor as part of a larger system. Such a network, or group of networks, could add significant computational capabilities to standard computers and associated interfaces.

For example, such a chip may be constructed utilizing a standard computer processor in parallel with a large physical neural network or group of physical neural networks. A program can then be written such that the standard computer teaches the neural network to read, or create an association between words, which is precisely the same sort of task in which neural networks can be implemented. Once the physical neural network is able to read, it can be taught for example to "surf" the Internet and find material of any particular nature. A search engine can then be developed that does not search the Internet by "keywords", but instead by meaning. This idea of an intelligent search engine has already been proposed for standard neural networks, but until now has been impractical because the network required was too big for a standard computer to simulate. The use of a physical neural network (i.e., physical neural network) as disclosed herein now makes a truly intelligent search engine possible.

A physical neural network can be utilized in other applications, such as, for example, speech recognition and synthesis, visual and image identification, management of distributed systems, self-driving cars and filtering. Such applications have to some extent already been accomplished with standard neural networks, but are generally limited in expense, practicality and not very adaptable once implemented. The use of a physical neural network can permit such applications to become more powerful and adaptable. Indeed, anything that requires a bit more "intelligence" could incorporate a physical neural network. One of the primary advantages of a physical neural network is that such a device and applications thereof can be very inexpensive to manufacture, even with present technology. The lithographic techniques required for fabricating the electrodes and channels therebetween has already been perfected and implemented in industry.

Most problems in which a neural network solution is implemented are complex adaptive problems, which change in time. An example is weather prediction. The usefulness of a physical neural network is that it could handle the enormous network needed for such computations and adapt itself in real-time. An example wherein a physical neural network (i.e., Knowm) can be particularly useful is the Personal Digital Assistant (PDA). PDA's are well known in the art. A physical neural network applied to a PDA device can be advantageous because the physical neural network can ideally function with a large network that could constantly adapt itself to the individual user without devouring too much computational time from the PDA. A physical neural network could also be implemented in many industrial applications, such as developing a real-time systems control to the manufacture of various components. This systems control can be adaptable and totally tailored to the particular application, as necessarily it must.

The embodiments and examples set forth herein are presented to best explain the present invention and its practical application and to thereby enable those skilled in the art to make and utilize the invention. Those skilled in the art, however, will recognize that the foregoing description and examples have been presented for the purpose of illustration and example only. Other variations and modifications of the present invention will be apparent to those of skill in the art, and it is the intent of the appended claims that such variations and modifications be covered. The description as set forth is not intended to be exhaustive or to limit the scope of the invention. Many modifications and variations are possible in light of the above teaching without departing from the scope of the following claims. It is contemplated that the use of the present invention can involve components having different characteristics. It is intended that the scope of the present invention be defined by the claims appended hereto, giving full cognizance to equivalents in all respects.

The embodiments of an invention in which an exclusive property or right is claimed are defined as follows:

1. A physical neural network based on nanotechnology, said physical neural network comprising:

at least one neuron-like node that sums at least one input signal and generates at least one output signal based on a threshold associated with said at least one input signal; and

at least one connection network associated with said at least one neuron-like node wherein said at least one connection network comprises a plurality of interconnected nanoconnections, such that each nanoconnection of said plurality of interconnected nanoconnections is strengthened or weakened according to an application of an electric field.

2. The physical neural network of claim **1** wherein said at least one output signal comprises a non-linear output signal based on said threshold.

3. The physical network of claim **1** wherein said at least one output signal comprises a linear output signal based on said threshold.

4. The physical neural network of claim **1** wherein said threshold comprises a threshold below which said at least one output signal is not generated and above which said at least one output signal is generated.

5. The physical neural network of claim **1** wherein said at least one connection network comprises:

a number of layers of said nanoconnections;

wherein said number of layers is equal to a number of desired outputs from said at least one connection network; and

wherein said nanoconnections are formed without influence by disturbances resulting from other nanoconnections thereof.

6. The physical neural network of claim 1 wherein at least one nanoconnection of said plurality of interconnected nanoconnections comprises an electrically conducting material.

7. The physical neural network of claim 6 wherein electrically conducting material is chosen such that a dipole is induced in said electrically conducting material in the presence of an electric field.

8. The physical neural network of claim 6 wherein said electrically conducting material comprises a chemically induced permanent dipole.

9. The physical neural network of claim 1 wherein said at least one nanoconnection comprises at least one nanoconductor.

10. The physical neural network of claim 9 wherein said at least one connection network comprises:

at least one connection network structure having a connection gap formed therein; a solution located within said connection gap;

wherein said solution comprises a solvent and said at least one nanoconductor; and

wherein an electric field applied across said connection gap to permit an alignment of at least one nanoconductor within said connection gap.

11. The physical neural network of claim 10 wherein said at is at least one nanoconductor is suspended in said solvent.

12. The physical neural network of claim 10 wherein said at least one nanoconductor is located at a bottom of said connection gap.

13. The physical neural network of claim 10 wherein an electrical conductance of said solution is less than an electrical conductance of said at least one nanoconductor within said solution.

14. The physical neural network of claim 10 wherein a viscosity of said solution permits said at least one nanoconductor to move when said electric field is applied across said connection gap.

15. The physical neural network of claim 10 wherein said at least one nanoconductor experiences an increased alignment in accordance with an increase in said electric field applied across said connection gap.

16. The physical neural network of claim 15 wherein nanoconnections of said at least one neuron-like node that are utilized most frequently by said at least one neuron-like node become stronger with each use thereof.

17. The physical neural network of claim 16 wherein said nanoconnections that are utilized least frequently become increasingly weak and eventually become unaligned.

18. The physical neural network of claim 16 wherein said at least one nanoconnection comprises a resistance, which is raised or lowered by a selective activation of said at least one nanoconnection.

19. The physical neural network of claim 9 wherein said at least one nanoconductor comprises a nanowire.

20. The physical neural network of claim 9 wherein said at least one nanoconductor comprises a nanotube.

21. The physical neural network of claim 9 wherein said at least one nanoconductor comprises a plurality of nanoparticles.

22. The physical neural network of claim 1 wherein at least one nanoconnection of said at least one connection network comprises a negative connection associated with said at least one neuron-like node.

23. The physical neural network of claim 22 wherein said at least one connection network comprises:

a number of layers of nanoconnections;

wherein said number of layers is equal to a number of desired outputs from said at least one connection network;

wherein said number of layers is equal to twice a number of desired outputs form said at least one connection network, if said nanoconnections comprise negative connections thereof; and

wherein said nanoconnections are formed without influence by disturbances resulting from other nanoconnections thereof.

24. A physical neural network apparatus based on nanotechnology, said physical neural apparatus comprising:

at least one connection network associated with at least one neuron-like node wherein said at least one connection network comprises a plurality of interconnected nanoconductors, such that each nanoconductor of said plurality of interconnected nanoconductors is strengthened or weakened according to an application of an electric field;

wherein each nanoconductor of said plurality of interconnected nanoconductors experiences an increase in alignment in accordance with an increase in said electric field;

wherein nanoconductors of said plurality of interconnected nanoconductors that are utilized most frequently by said at least one neuron-like node became stronger with each use thereof; and

wherein nanoconductors of said plurality of interconnected nanoconductors that are utilized least frequently become increasingly weak and eventually become unaligned.

25. The physical neural network apparatus of claim 24 wherein said at least one neuron-like node sums at least one input signal and generates at least one output signal based on a threshold associated with said at least one input signal.

26. The physical neural network of claim 24 wherein said at least one connection network comprises a plurality of layers, wherein a number of layers of said plurality of layers is equal to a desired number of outputs from said at least one connection network.

27. A method for assembling a physical neural network based on nanotechnology, said method comprising the steps of:

forming at least one neuron-like node that sums at least one input signal and generates at least one output signal based on a threshold associated with said at least one input signal; and

configuring at least one connection network associated with said at least one neuron-like node wherein said at least one connection network comprises a plurality of interconnected nanoconnections, such that each nanoconnection of said plurality of interconnected nanoconnections is strengthened or weakened according to an application of an electric field.

28. The method of claim 27 wherein said at least one output signal comprises a non-linear output signal based on said threshold.

29. The method of claim 27 wherein said at least one output signal comprises a linear output signal based on said threshold.

30. The method of claim 27 wherein said threshold comprises a threshold below which said at least one output signal is not generated and above which said at least one output signal is generated.

31. The method of claim 27 further comprising the steps of:

configuring said at least one connection network to comprise a number of layers of said nanoconnections, wherein said number of layers is equal to a number of desired outputs from said at least one connection network; and

forming said nanoconnections without influence by disturbances resulting from other nanoconnections thereof.

32. The method of claim **27** comprising the step of; forming at least one nanoconnection of said plurality of interconnected nanoconnections from an electrically conducting material.

33. The method of claim **32** further comprising the step of: configuring said electrically conducting material such that a dipole is induced in said electrically conducting material in the presence of an electric field.

34. The method of claim **32** comprising the step of: chemically inducing a permanent dipole within said electrically conducting material.

35. The method of claim **27** wherein said at least one nanoconnection comprises at least one nanoconductor.

36. The method of claim **35** further comprising the steps of:

forming a connection gap from at least one connection network structure associated with said at least one connection network;

locating a solution within said connection gap; configuring said solution to comprises a solvent and said at least one nanoconductor; and

applying an electric field across said connection gap to permit an alignment of at least one nanoconductor within said connection gap.

37. The method of claim **36** further comprising the step of: suspending said at least one nanoconductor in said solvent.

38. The method of claim **36** further comprising the step of: locating said at least one nanoconductor at a bottom of said connection gap.

39. The method of claim **36** further comprising the step of: configuring said solution such that an electrical conductance of said solution is less than an electrical conductance of said at least one nanoconductor within said solution.

40. The method of claim **36** further comprising the step of: configuring said solution to comprise a viscosity that permits said at least one nanoconductor to move when said electric field is applied across said connection gap.

41. The method of claim **36** further comprising the step of: increasing an application of said electric field across said connection gap to thereby permit said at least one nanoconductor to experience an increased alignment in accordance with an increase in said electric field applied across said connection gap.

42. The method of claim **41** wherein nanoconnections of said at least one neuron-like node that are utilized most frequently by said at least one neuron-like node become stronger with each use thereof.

43. The method of claim **42** wherein said nanoconnections that are utilized least frequently become increasingly weak and eventually become unaligned.

44. The method of claim **42** wherein said at least one nanoconnection comprises a resistance, which is raised or lowered by a selective activation of said at least one nanoconnection.

45. The method of claim **35** further comprising the step of: configuring said at least one nanoconductor to comprise a nanowire.

46. The method of claim **35** further comprising the step of: configuring said at least one nanoconductor to comprise a nanotube.

47. The method of claim **35** further comprising the step of: configuring said at least one nanoconductor to comprise a plurality of nanoparticles.

48. The method of claim **27** wherein at least one nanoconnection of said at least one connection network comprises a negative connection associated with said at least one neuron-like node.

49. The method of claim **48** further comprising the step of:

configuring said at least one connection network to comprise a number of layers of nanoconnections, wherein said number of layers is equal to a number of desired outputs from said at least one connection network;

wherein said number of layers is equal to twice a number of desired outputs form said at least one connection network, if said nanoconnections comprise negative connections thereof; and wherein said nanoconnections are formed without influence by disturbances resulting from other nanoconnections thereof.

50. A method for assembling a physical neural network apparatus based on nanotechnology, said method comprising the steps of:

forming at least one connection network associated with at least one neuron-like node wherein said at least one connection network comprises a plurality of interconnected nanoconductors, such that each nanoconductor of said plurality of interconnected nanoconductors is strengthened or weakened according to an application of an electric field;

wherein each nanoconductor of said plurality of interconnected nanoconductors experiences an increase in alignment in accordance with an increase in said electric field; wherein nanoconductors of said plurality of interconnected nanoconductors that are utilized most frequently by said at least one neuron-like node become stronger with each use thereof; and

wherein nanoconductors of said plurality of interconnected nanoconductors that are utilized least frequently become increasingly weak and eventually become unaligned.

51. The method apparatus of claim **50** further comprising the steps of: summing at least one input signal via said at least one neuron-like node; and generating at least one output signal based on a threshold associated with said at least one input signal.

52. The method of claim **50** further comprising the step of: configuring said at least one connection network to comprise a plurality of layers, wherein a number of layers of said plurality of layers is equal to a desired number of outputs from said at least one connection network.

* * * * *